

A challenge to genetic transparency

Those engaged in the publicly funded effort to sequence the human genome should look on their new rival as healthy competition. But they will also need to protect the standards they have fought hard to establish.

Two weeks ago, when the laboratory equipment company Perkin-Elmer and genomics pioneer J. Craig Venter unexpectedly announced that they were joining forces to carry out a privately funded sequencing of the complete human genome, they sent tremors through the sequencing community. The language in which the announcement was made, speaking of a "breakthrough" in DNA analysis technologies that would allow an "ultra-high" throughput of samples, only reinforced concern. If it was really true that, as some press reports claimed, the new initiative would achieve its objective several years sooner, and ten times more cheaply, than publicly funded efforts, how much longer would it be before public agencies decided to pull the plug and divert funds elsewhere?

Criticisms

By the end of last week, much of the initial alarm had dissipated. Closer inspection of Perkin-Elmer's plans for its new sequencing machine revealed that, while a substantial improvement on current technology, it does not represent a quantum advance over that being offered by rival companies — while many of its claims to speed and effective automation remain to be demonstrated. Leaders of the government-funded efforts, such as Francis Collins, director of the US National Centre for Human Genome Research, made much of the alleged shortcomings in Venter's 'shotgun' approach to sequencing to argue the continued need for their own, more accurate, sequencing endeavours.

Such criticisms may be too defensive. But it is indeed important that there should be a reaffirmation of the publicly funded programme. In the Congress, staff on the committees responsible for the budget of the National Institutes of Health (NIH) have rightly made reassuring noises about the need to ensure that the \$1.9 billion of public funds already invested in the Human Genome Project do not go to waste, or merely benefit the private sector (see page 201). Britain's Wellcome Trust last week gave public-sector sequencers a shot in the arm by its fortuitously coincidental decision to double its support for human genome sequencing, contributing a further £110 million (\$175 million) over seven years — a move allowing it to shoulder responsibility for one-third of the total effort. The talk was of healthy competition rather than throwing in the towel.

Those attending last week's genomics meeting at Cold Spring Harbor, who had lived through an emotional week as details of the Venter/Perkin-Elmer initiative became known, greeted the Wellcome announcement, and the firm commitment to open access to sequencing data that accompanied it, with a visible sense of enthusiasm and relief. But it would be unwise for researchers to underestimate the magnitude of the technical and scientific challenge that the new project will present. Nor should they overestimate the depth of support in a Congress that frequently expresses its reluctance to pick up the bill for activities able to find financing from the private sector.

On the technical aspects, the public-sector sequencers have

powerful benefits on their side. Public funding (including that from Wellcome) has ensured that a concern for accuracy and completeness has not been dominated by excessive demands for speed or high returns on investment. Collins has established high standards of quality that the half-dozen federally funded US sequencing laboratories are expected to achieve with their data, obtained by meticulously working through ordered fragments of DNA. Venter's strategy, although faster and probably cheaper, cannot yet claim to compete on the contiguity of the final product.

Challenges

The technical challenges arising from the initiative represent good and bad news. Positively, speeding up of sequencing will greatly increase the volume of material requiring integration into public databases and analysis. On the other hand, Venter's stated plans make it clear that the sequences will not be complete. Those interested commercially in sequencing primarily as a source of potential new diagnostic techniques and therapies can be depended on to pursue the level of accuracy required to understand the detailed functioning of individual genes. But they will have little interest in completeness for its own sake.

Prompt availability is an even more serious worry. Venter and Perkin-Elmer have promised to make their sequence data freely available to researchers, but they will be released only at three-monthly intervals — and then only in the form of 'consensus' sequences. Researchers will be denied access to the raw data from which these sequences are calculated, and thus unable to check their accuracy or to integrate the data into the continuing public sequencing programme. Perkin-Elmer has already made it clear that its interest in bankrolling the new sequencing initiative is not only to use it as a test-bed for sequencing machines that will then be sold to others, but also to profit from the data — in particular information about genetic diversity, or 'polymorphisms', that are likely to emerge. Where fast access to raw data conflicts with pressure to skim it for potential commercial value, it is not difficult to see where the balance in a commercial venture is likely to be struck. That, in the absence of a firm commitment to public funding for sequencing, would be a significant threat to the science.

So far, those responsible for carrying out the Human Genome Project on both sides of the Atlantic have shown themselves impressively aware of the need for continual vigilance to ensure that commercial pressures are not allowed to encroach excessively. The Venter/Perkin-Elmer initiative must be welcomed as an important competitor in the genome sequencing stakes; even the most dedicated public enterprise benefits from healthy private competition. Arguments about ownership are inevitable. But for aggressive scientific entrepreneurs, collaboration with the publicly funded community is more profitable than hostility. That is a card the community has up its sleeve and should be prepared to play in maintaining essential values of complete scientific access to our genetic inheritance. □



India's nuclear tests meet with domestic praise and protests

[NEW DELHI] Exercising the nuclear option seems to be one of the few election promises that India's Bharatiya Janata Party (BJP) government has so far fulfilled (see *Nature* 392, 320; 1998). While prime minister Atal Behari Vajpayee claims they were carried out in the interest of national security, the tests last week on Buddha's birthday in the land of Mahatma Gandhi have blasted a large hole in India's image as a spiritual, non-violent country.

"India has shot itself in the head," says Proful Bidwai, an anti-nuclear activist. While BJP celebrated the nuclear test with fire crackers, some 500 intellectuals and others belonging to nongovernmental organizations held a protest march in the capital with slogans: "We want water, bread; not bombs."

Prominent critics of the test included Kuldip Nayar, the former high commissioner to the United Kingdom. India's *Economic Times* newspaper said in an editorial: "For a country with a third of its population below the poverty line, spending its scarce resources on a nuclear arsenal could be damaging," adding that after the test, "India's strategic position has not improved and may have deteriorated".

Although critics call the tests a political diversion, a polling of 1,007 adults in six cities showed that 91 per cent approved of the tests. Reaction in the scientific community at large

has also been supportive of BJP's action.

When the science minister, Murli Manohar Joshi, broke the news at a conference of directors of the Council for Scientific and Industrial Research — India's largest scientific agency — all 40 assembled directors hugged each other and applauded. The head of the council, Ragunath Mashelkar, declared that he was "extremely proud as an Indian".

The former chairman of the Atomic Energy Commission, Raja Ramanna, is reported to have opened a bottle of champagne to celebrate, while Udippi Rama Rao, former chairman of the space commission (now chairman of the UN Committee on Outer Space) and Govindarajan Padmanabhan, director of the Indian Institute of Science, both said that India should have carried out the tests a long time ago.

Krishnagopala Iyengar, another ex-chairman of the Atomic Energy Commission, congratulated Vajpayee "for having erased doubts over the competence of Indian scientists". He called for research on fourth-generation bombs "based on nuclear isomers and super-heavy elements".

"We don't have even a minority of scientists who think about ethics because most scientists are government servants," complains Bidwai.

Surprisingly, M. G. K. Menon, former



Broken image: Shakti 2 in India's Thar desert, one of the four nuclear test sites.

president of the International Council of Scientific Unions, who joined the BJP on the eve of the elections (see *Nature* 391, 627; 1998), seems to have played no role in the nuclear testing.

And, although scientists have been praising the testing in their individual capacities, no professional science body has officially

Nuclear tests were a culmination of twenty-five years of planning

[NEW DELHI] India, which surprised the world last week by carrying out underground nuclear tests, has been secretly working on the design of a whole range of nuclear weapons for the past 25 years.

The weapons programme had in fact remained under military control all along, something India managed to keep under wraps, in the same way that it was able to hide preparations for the latest tests — which included explosion of a 45-kilotonne hydrogen bomb at Pokran in Rajasthan — from US spy satellites.

The revelations emerged on 17 May during a joint press conference by Rajagopalan Chidambaram, the Department of Atomic Energy (DAE) secretary, and A. P. J. Abdul Kalam, chief of

the Defence Research Development Organization (DRDO), at which they showed a video and gave some details of the underground tests — code-named Shakti-98 — carried out on 11 and 13 May (see over).

"Based on 25 years of research and development, we have designed and developed various kinds of nuclear explosives — fission, boosted fission, thermonuclear and low-yield," said Chidambaram. All except the boosted fission bomb were tested last week.

Domestically produced enriched uranium-235 is believed to have been used in some of these bombs. But Chidambaram said that this information was classified. "All I can say is that the fissile material used was totally indigenous."

Kalam revealed that the devices tested were the result of a "decades-old" partnership between DRDO and DAE. "When nuclear technology and defence technology met, they got transformed to nuclear-weapons technology."

DRDO, he said, had been involved in the design and manufacture of detonators, electronic trigger systems and chemical explosives, besides making contributions "in aerodynamics, arming, fusing and safety interlocks".

According to Kalam, those tested last week were not experimental devices, but deployable, computer-designed and certified bombs. The yield of the H-bomb — set off at a depth of 300 metres — was deliberately designed to be low so as to avoid

damage to the closest village, 5 km away from ground zero.

"The tests marked the culmination of militarization which started years ago," said Kalam. He added that the nuclear warheads now available can be put into the domestically produced Prithvi surface-to-surface missile and the intermediate-range Agni missile, which is ready for mass production and whose newer longer-range version is at an "advanced stage of development at DRDO".

This is the first admission that India's nuclear weapons programme not only continued under different governments after the 'peaceful nuclear explosion' on 18 May 1974, but was actually militarized long before the Bharatiya Janata Party came to power two months ago. **K.S.J.**

endorsed the nuclear adventure. "This is a political and defence matter on which our academy cannot take any stand," says Srinivasan Varadarajan, president of the Indian National Science Academy in New Delhi.

Varadarajan also doubts that the latest scientific feat is likely to create a climate for more funding for scientific research in general. "More money will certainly start flowing to defence and atomic research, but these are areas that are already getting the biggest share of our budget."

Sanctions by the United States and other countries would admittedly hurt several health and development projects. The effects of the tests on India's science establishments have yet to be gauged, but no one seems particularly worried. "Sanctions are welcome," says Mashelkar, pointing out that India "seems to excel precisely in those areas in which technology is denied".

The Indian Space Research Organization, which has learnt to live with a variety of sanctions in the past, says "sanctions will pinch, but not hurt". But some defence projects, such as the light combat aircraft which critically depends on a flight control system from Lockheed Martin of the United States, may be in jeopardy.

Work on the main battle tank, which uses German engines, and the advanced light helicopter, which requires components from the United States, may also be affected. But Kalam says denials will help spur self-reliance. "No one can throttle us."



Approval from K. Santhanam, R. Chidambaram and Abdul Kalam of the atomic energy establishment.

Gyanendra Nath, however, an adviser in the international division of the Ministry of Science and Technology, says that repercussions in many collaborative projects "are going to be serious". He says cancellation of DM300 million (US\$168.3 million) in grants to India will jeopardize a two-year Indo-German programme on nanotechnology.

Germany has also been funding research in superconductivity at the National Physical Laboratory in New Delhi and is a major donor (\$20 million) to the Indian Institute

of Technology in Madras. With Japan cutting off its aid, the Indo-Japanese project at the SPring-8 synchrotron may close down, says Nath.

Since the winding up of the India-US fund in January this year, there has been no major US-funded project in India, and officials say the United States cannot afford to take actions that will hurt collaborative projects in weather or vaccine research "since they [the United States] will be the losers."

K.S. Jayaraman

Disappointment as New Zealand's budget gives science 'inflation'

[CANBERRA] Research gained a small increase in the budget of the New Zealand government, delivered last week. But the increase received a mixed reception, and New Zealand science seems to be following other downward trends evident in the Australian budget revealed two days earlier (see *Nature* 393, 101; 1998).

Maurice Williamson, New Zealand's minister for research, science and technology, announced that NZ\$10 million (US\$5.34 million) more would be allocated to research in all portfolios, bringing the 'science envelope' to a total of NZ\$600 million, and representing an increase roughly in line with expected inflation.

Of this extra money, the Public Good Science Fund (PGSF), administered by the Foundation for Research Science and Technology, will receive another NZ\$9 million, bringing its funding to NZ\$317.3 million for 1998-99. The remaining NZ\$1 million is directed to "non-specific output funding" for nine government-owned Crown Research Institutes.

Science leaders have praised Williamson for holding the line against reported cabinet opposition and keeping spending on research

constant in real terms. Yet there is a marked shortfall from the extra NZ\$25 million that Williamson had earlier stated was needed to keep up with the growth in gross domestic product (GDP) (see *Nature* 391, 426; 1998).

In a statement, Williamson claimed that the increase in dollars is "a continuation of the government's commitment to increase public investment to 0.8 per cent of GDP by 2010". But observers predict that the national expenditure on research and development relative to GDP will slip below the last stated New Zealand figure of 0.52 for 1995-96.

The PGSF is the principal source of competitive funding for Crown Research Institutes, as well as for projects by researchers in the country's seven universities. An extra NZ\$7.7 million is being directed by Williamson to 17 areas of the fund in proportions largely unchanged from the past two years.

The 18th area, the Health Research Council, will receive an extra NZ\$1.5 million, taking its total to NZ\$33.2 million. But this research council has also been moved from marginal to full-cost funding of research grants within three years, and the result is likely to be a fall in the size of grants.

The Marsden Fund for competitive grants in basic research, run by the Royal Society of New Zealand, will remain at last year's level of NZ\$22 million, and the government has dropped its own target of increasing the fund to 10 per cent of the PGSF. The president of the academy, George Petersen, speaking in a personal capacity, "particularly regrets" this decision.

Although acknowledging fiscal restraints, he says it "will require massive increases to bring spending on basic research back on track. The budget sends very poor signals to the scientific community and the government is clearly putting other priorities ahead of science."

Before the publication of a government white paper detailing decisions on a review of higher education (see *Nature* 392, 320; 1998), the education minister, Wyatt Creech, announced a massive change in the university funding system to a new taxpayer-funded 'Universal Tertiary Tuition Allowance'. While vice-chancellors welcome the change, staff and students are violently opposing universities being driven by students' choices through portable, personal 'vouchers'.

Peter Pockley

First underground nuclear waste store set to open in US

[WASHINGTON] The US Department of Energy is on the brink of delivering its first consignment of radioactive waste to permanent storage. This follows certification by the Environmental Protection Agency (EPA) that the Waste Isolation Pilot Plant (WIPP) near Carlsbad, New Mexico, is ready to receive it.

Unless environmentalists succeed in two last-ditch legal challenges, the first shipment of low-level transuranic waste will arrive at WIPP from the Los Alamos National Laboratory on or soon after 19 June. The waste is clothing and equipment that has been contaminated by plutonium or other actinides in the US nuclear weapons complex, and which is now stored in steel barrels at 23 locations.

Although the waste is far less radioactive than the spent nuclear fuel and other highly radioactive waste that the United States hopes eventually to store at Yucca Mountain, Nevada, the certification of WIPP was described by Federico Peña, the energy secretary, as a "historic milestone".

The opening of WIPP, in a salt bed 2,000 feet underground, would mark the first time the United States has buried any nuclear waste in a repository that is supposed to be permanent — and the first time in the world that such waste has been deposited deep underground. Sweden and Finland have permanent repositories, but these are close to the land surface.

Warren Weart of the Sandia National Laboratory, New Mexico, a senior manager on the WIPP project for 25 years, said last week that he is not taking a successful operation for granted. "I'll hold off on a final celebration until we actually place some waste," he says. "I will be surprised if [opponents] don't try to block it or slow it down."

Don Hancock, director of the Southwest Research and Information Center in Albuquerque, New Mexico, says that two lawsuits attempting to block the opening are pending — one against the EPA and one against the Department of Energy. They may attract the support of Tom Udall, the attorney-general of New Mexico. In contrast to Yucca Mountain, WIPP is supported by the governor and both senators in its home state.

Although not very radioactive — a tonne of it contains less than one-tenth of a curie — the transuranic waste contains elements such as plutonium which have long half-lives. So the argument about WIPP has focused on its ability to maintain its integrity for thousands of years. In 1996, a National Academy of Sciences panel concluded that it could do so if undisturbed by humans — but that a lack of disturbance, by oil drilling for example, could not be guaranteed. **C.M.**

Clinton 'is failing to honour pledge on AIDS vaccine'

[WASHINGTON] President Bill Clinton has failed to do enough to accelerate AIDS vaccine development, a year after pledging to develop an AIDS vaccine within a decade, according to US vaccine research advocates.

A critical report issued last week by the AIDS Vaccine Advocacy Coalition (AVAC), a pressure group based in San Francisco, complains that the National Institutes of Health (NIH) has failed to appoint a director for the new vaccine research centre that Clinton set up to lead the effort (see *Nature* 387, 323; 1997). It also says that the president has done nothing in the interim to follow through on his pledge. "A few inspiring speeches do not constitute leadership," the group says.

Some of AVAC's criticisms are echoed by the International AIDS Vaccine Initiative, an operation supported by the World Bank and the United Nations AIDS programme. "We believe the president can do more," says Seth Berkley, president of the initiative, who says the ten-year goal will be met only if current efforts are accelerated.

AVAC says that NIH has faced a struggle in hiring a director for its new vaccine research centre because the individual appointed would have no autonomous budget, and would lack a well-defined place in the NIH management hierarchy.

Anne Thomas, a spokeswoman for NIH, says that the appointment is subject to "considerations that are as important as speed",

adding that "the search is active and we hope to have someone on campus this year". Of the general criticism, she says: "We think the work is progressing well in the vaccine programme, both intramural and extramural."

AVAC also says that corporations are not doing enough vaccine research. It singles out the pharmaceutical company Smith-Kline Beecham which, it says, has \$1.2 billion in annual vaccine revenues but "essentially no active HIV-vaccine development programme".

But the group does applaud the speed with which the National Institute of Allergy and Infectious Diseases issued 58 new "innovation grants" for vaccine research last year, and says the share of AIDS research funding spent on vaccines at NIH has grown from 6.7 per cent in 1995 to 9.5 per cent this year.

The comments of the two groups reflect long-standing concerns on the part of some physicians and AIDS activists that an unbridgeable gap still divides basic research from the trial and development of an AIDS vaccine. Many researchers, however, consider this gap less of a problem than the dearth of promising vaccine candidates for trial.

Scott Hitt, a Los Angeles physician who chairs the President's Advisory Committee on HIV/AIDS, says: "The president has stated a very ambitious goal. Until the vaccine is out, there will always be criticism that we're not doing enough." **Colin Macilwain**

Japan's emissions bill comes under fire

[TOKYO] Proposed Japanese legislation to reduce greenhouse-gas emissions has come under criticism for lacking measures necessary to ensure compliance by industry.

The bill, drawn up by the Environment Agency and now under consideration in parliament, provides a basic framework for Japan's emissions cuts. It is thought to be the first of its kind to be submitted by one of the signatory nations to the United Nations Framework Convention on Climate Change after last December's meeting in Kyoto.

The bill would require both the central and prefectural governments to draw up plans to reduce greenhouse-gas emissions, and to make frequent reports on their progress. It does not, however, require industry — the largest emitter of greenhouse gases — either to disclose emission levels or to draw up plans for effective reduction of such gases.

Hiroshi Oki, Japan's environment minister and chairman of the Kyoto conference, said in a statement last week that the proposed legislation "cannot be

rated 100 per cent". But it nevertheless "contains measures which are realistic at present", he said.

Environment agency officials insist that the significance of the bill lies in its suggestion that companies should adopt 'self-regulatory measures'. But critics such as members of the Kiko Forum, a federation of environmental groups, argue that Japan is unlikely to achieve its target — a six per cent cut from 1990 levels — by relying on industry's goodwill.

The preliminary draft of the bill, which would have imposed stricter measures on industry including mandatory reporting of plans for reduction of emissions, was heavily revised after it met resistance from the Ministry of International Trade and Industry (MITI) and industrial lobbyists. MITI says energy conservation legislation already provides measures for emissions reduction. But the energy-reduction bill does not specify measures to be taken on greenhouse gases, and does not apply to all of Japan's industry. **Asako Saegusa**

Europe's life patent moratorium may go ...

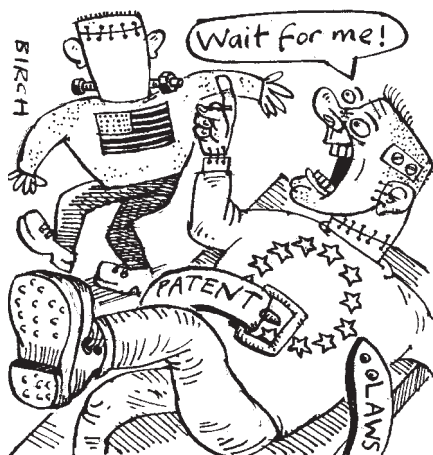
[MUNICH] A moratorium imposed by the European Patent Office (EPO) in 1995 on the patenting of plants and animals is likely to be lifted following last week's approval by the European Parliament of a new European Union (EU) directive on biotechnological inventions.

The directive will also speed up patent procedures because it makes clear what is, and what is not, patentable. But the moratorium may not be lifted for another year.

Officials of the EPO, which is not an EU body, have welcomed the reference point that the directive will provide for its own rules, which were written into its European Patent Convention (EPC) in 1973, before the advent of recombinant DNA technology. These rules, whose interpretation is determined by case law, have proved ambiguous in relation to biotechnological inventions.

According to the new directive, elements of the human body — including partial sequences of genes — are not patentable; nor are procedures for human cloning. But isolated human genes can be patented, provided that their function is known and an application defined, for procedures such as diagnostics.

Transgenic animals or plants are also patentable. Parliament rejected the draft directive the first time round, in March 1995,



partly because of ethical concerns about these issues.

Groups opposed on principle to the patenting of organisms created by genetic engineering routinely challenge all such patent applications, citing a clause in the EPC which states that a patent should not be granted if it is "contrary to public order or morality". EPO boards of appeal have always rejected these challenges.

But opponents have had more luck with a semantic argument relating to another EPC clause, originally intended to protect the rights of plant breeders, which excludes plant and animal varieties from patentability.

This clause was never intended to forbid the patenting of plants and animals in general. But in February 1995, an EPO board of appeal accepted opponents' arguments that the term 'plant' can be considered an umbrella term for a collection of plant varieties — and is thus not patentable under the terms of the convention (see *Nature* 374, 8; 1995).

Since then, the EPO has issued no patents on animals or plants. Hundreds of applications are therefore pending, awaiting further clarification of the principle.

The EPO hopes that the 1995 precedent will be reversed by its Enlarged Board of Appeals — the EPO's highest legal authority — which has just been asked to interpret the question of plant patentability following a challenge to a Novartis patent on a plant genetically engineered to be herbicide-resistant.

EPO officials decline to anticipate the outcome of the Enlarged Board of Appeals. But the board is now almost certain to take the new directive as its guidance, particularly as its formal request for clarification was delayed until after the parliamentary vote.

In the unlikely event that it rejects the directive as guidance, then, says Christian Guggerell, an EPO expert on biotechnological inventions, the 18 signatories of the EPO will have to agree to rewrite the EPC rules, a process that could take many years.

Simon Cohen, a London-based patent lawyer with the firm of Taylor Joynson Garrett, says that, as the directive makes clear that human genes with a described function and application as well as transgenic animals and plants are patentable, there will be fewer oppositions to patents on biotechnological inventions. "This will speed up patenting procedures for biotechnology companies."

He points out, however, that there will still be plenty of room for challenges based on ethical considerations, as the directive excludes from patentability "processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal".

Brian Yorke, head of corporate intellectual property at Novartis, agrees that the "clarity and balance" of the patent directive will make life much easier for industry. He believes that the Novartis case being considered by the EPO Enlarged Board of Appeal can — and probably will — be interpreted "as being in line with the EU patent directive". Otherwise, he says, the EPO would be out of line with future national legislation in EU countries, all of which are EPO member states.

The directive on biotechnological inventions, whose formal approval is assured, must be incorporated into national law within two years by all EU member states. **Alison Abbott**

... as US office claims right to rule on morality

[WASHINGTON] The US patent office is claiming it has the authority to stop the issuing of a patent, on moral grounds alone, for ways to make human-animal chimaeras (see *Nature* 392, 423, 1998). The office is currently faced with a broad-ranging application for such a patent.

But the applicants — Jeremy Rifkin, the president of the Foundation on Economic Trends, a Washington advocacy group, and Stuart Newman, a developmental biologist at New York Medical College in Valhalla, New York — say that there is no legal basis for such an assertion.

Even if there was such a basis, they say, it would be impossible for the Patent and Trademark Office (PTO) to draw a line defining what is and what is not morally acceptable.

Rifkin and Newman filed for a patent last December on methods for creating a 'human/animal chimaera', and say that their intention is to raise a broad public and legal debate about the implications of such a patent.

On the same day (2 April) that news first broke of the application, Bruce Lehman, the PTO commissioner, issued a statement asserting that the PTO's legal authority allows it to deny patents deemed to be "injurious to the well-being, good policy or good morals of society".

Lehman's statement said it was the position of the PTO that inventions directed to human/non-human chimaeras could not be patentable "because, among other things, they would fail to meet the public policy and morality aspects" of patent law". Lehman says that this position is based on a court

decision of 1817.

Lehman also argued that Rifkin and Newman are engaged in an attempt "to panic people into an overreaction". But the two applicants say that their application is based on perfectly feasible technology, and that patent law gives Lehman no 'moral' grounds for blocking an application.

Even if it did, they say, it would be impossible for the PTO to draw a line between 'moral' human-animal mixes and 'monsters'. Newman asks: what would rationally distinguish their creation from inventions that have already been applied for, such as pigs carrying human genes as sources of transplantable organs?

Rifkin has already said that he would mount a legal challenge to any rejection of his application by the patent office. **Meredith Wadman**

Genome effort 'still in need of support'

[WASHINGTON] The leaders of the publicly funded Human Genome Project took a firm stance last week to challenge any suggestion that the announcement of a parallel, privately funded initiative made their efforts redundant.

The challenge has come from a plan by the genome researcher J. Craig Venter, president of The Institute for Genomic Research in Rockville, Maryland, and the company Perkin-Elmer to launch a new venture.

This will be based around a novel analyser that the company has developed, and Venter's alternative 'shot-gun' sequencing strategy, to carry out an independent sequencing of the complete human genome (see *Nature* 393, 101; 1998).

The announcement has stimulated widespread speculation that the Human Genome Project, to which the US government has promised \$3 billion over 15 years, starting in 1990, may no longer be necessary.

But in a letter to *The New York Times*, written in response to a suggestion that the project might be redirected, Harold Varmus, the director of the US National Institutes of Health, sharply disputed any such conclusion. "This is not the case," he wrote.

"The ambitious milestones set by the Human Genome Project since it began have thus far been reached." Varmus pointed out that data produced by the project have made the private effort possible and have already reduced the time needed to isolate disease-related genes from decades to often just days.

Much of the scepticism about whether a major federal project is still needed has come from within the biotechnology community. Messages — usually anonymous — appearing on industry web-sites have contained suggestions that, for example, "it would be a much better use of taxpayers' money not to go to government-supported human genome sequencing".

But, within the academic research community, the feeling remains strong that, whatever sequence emerges from the new initiative, it cannot match the completeness — and hence full potential value — of the efforts being funded by the US and other governments, as well as Britain's Wellcome Trust (see right).

"What the new company is proposing to do is an incomplete job," says John Sulston, director of the Sanger Centre, outside Cambridge, England. "We know that the sequence it plans to produce will have gaps in it; they may be deferring some of the costs involved in obtaining a complete sequence, but the result is that what they will produce has less power of interpretation."

Sulston and others also say that they are not convinced that the conditions under which Venter has promised to make his data



Varmus: genome project is on target.

publicly available to other researchers will be sufficient to meet their needs.

Such arguments are already being strongly deployed by those seeking to ensure that the US Congress maintains its commitment to funding the genome project, to which it has agreed to allocate \$170 million in the current fiscal year, out of a total allocation to the National Human Genome Research Institute of \$217 million.

So far, the message seems to be getting through. "I don't think that there are many people in Congress who are looking at this and saying, 'Gee, this is great. The private sector is going to do all the work on unlocking the genetic code. Let's move on,'" says Dave Kohn, a spokesman for John Porter (Republican, Illinois), the chairman of the

Labor/Health and Human Services/Education subcommittee of the House of Representatives Appropriations Committee.

The subcommittee writes the spending bill that funds the National Institutes of Health. Kohn says that, in his discussions with Porter and other subcommittee members, support for the Human Genome Project remains "very, very strong. There continues to be a great deal of interest in having federally supported research in this area."

Even some industry executives remain confident that the government's efforts will remain predominant. "My bet would be that [the Human Genome Project] would be the ultimate provider of the polished, sequenced human genome," says Roy Whitfield, chief executive officer of Incyte Pharmaceuticals of Palo Alto, California.

Incyte is itself deeply involved in sequencing efforts. Whitfield says: "I would hope that [publicly funded researchers] would continue with their efforts, or even redouble them."

David Dickson & Meredith Wadman

British funding boost is Wellcome news

[LONDON] When Michael Morgan, programme director of genetics at Britain's Wellcome Trust, addressed a genomics meeting at Cold Spring Harbor, New York, last Friday (15 May) on the trust's plans for supporting human sequencing efforts, he received almost a standing ovation from the researchers present.

Morgan explained a decision by the trust the previous weekend to more than double its commitment to the Human Genome Project, bringing its total investment to £205 million (US\$340 million).

The extra money will allow the Sanger Centre near Cambridge, England, which is jointly run with the UK Medical Research Council, to undertake the sequencing of one-third of the complete genome; the trust had previously been committed to financing one-sixth.

But the trust has indicated that it is prepared to go even further. Morgan said it intends to open discussions with existing members of the project — most of whom are

based in the United States — under which up to half of the genome could be sequenced in Britain.

Although Wellcome's decision to increase funding was taken independently of a privately funded sequencing initiative by J. Craig Venter and Perkin-Elmer (see above), its announcement is said to have been brought forward to provide a riposte to some of the claims being made by the latter.

Even more than the vote of confidence in the genome project that the trust's decision represents, it was Morgan's statement of the firm position the trust is taking on the question of access to sequencing data that brought the researchers to their feet.

"What Wellcome has done is fabulous," says Richard Gibbs of the Baylor University College of Medicine, a co-organizer of the Cold Spring Harbor meeting. "The warm reaction to Morgan's presentation reflected the impassioned feeling among those present

that the data needs to be freely and publicly available."

Most of those at the meeting are involved in publicly funded sequencing efforts. Surveys have revealed deep-rooted concern in this community about the impact of the encroachment of patent claims on genetic sequences (see *Nature* 392, 325; 1998). So their support for Wellcome's position came as little surprise.

This support extended to Morgan's announcement that the trust is not only conducting an urgent review of the credibility and scope of patents based solely on DNA sequences but is also prepared, where appropriate, to challenge such patents.

Wellcome's initiative appears to have helped to restore confidence in the public sequencing effort at a critical time. Morgan admits that this has been part of his goal, pointing to a similar catalytic effect which he claims the trust was able to exercise on European-wide efforts to sequence the yeast genome.

D. D.

UN reaches deal on biodiversity science

[LONDON] A meeting of signatory states to the United Nations biodiversity convention ended in Bratislava, Slovakia, last week with agreement on a compromise over whether the convention should be provided with independent scientific advice.

The meeting agreed to set up expert panels by 2000 on marine biodiversity, and inland water biodiversity. A working group to establish guidelines on bioprospecting will be set up immediately (see *Nature* 392, 535–540; 1998).

But the panels will not be totally independent. The bioprospecting group will be appointed directly by governments. Members of the other expert panels will be picked by the convention's executive secretary from shortlists from each of the convention's five regions: Africa, Asia, Latin America, Eastern Europe and Western Europe.

The convention's 'parties' also agreed to strengthen its formal body of government-appointed scientists, which will meet twice before the next biodiversity conference in 2000, instead of just once. And the Australian government agreed to help find ways of increasing the availability of taxonomy expertise, particularly in developing countries.

The convention's budget was increased by a modest 7.5 per cent to US\$8.6 million.

And a deadline of February 1999 was set to finalize the protocol regulating the safety of genetically modified organisms.

The agreement on expert panels amounts to a compromise between developed and advanced developing country parties on the one hand, and least developing country parties on the other.

The former had proposed setting up panels of scientists from outside the convention to advise it on scientific issues related to biodiversity. These governments favoured such independent expertise on the grounds that the convention's existing advisory body met too infrequently.

They also argued that this body is unable to offer practical advice, given that the views of its member scientists tend to reflect the policies of their governments.

But decisions at UN conventions need a consensus of all the parties. And the least developed countries had signalled their opposition to extra scientific advice in the conference's first week (see *Nature* 393, 99; 1998). They were concerned that they would not have sufficient control over the new science panels, particularly if they were dominated by scientists from developed countries.

The decision to set up a group to formulate guidelines on bioprospecting stems from

continued controversy over access to genetic resources. It will be welcomed, in particular, by pharmaceutical companies, which are increasingly concerned at what they perceive to be a trend in developing countries to pass tough laws regulating such companies, and other research organizations, when prospecting for natural products.

Under a Philippines law, for example, all scientists — whether local or from abroad — must seek permission directly from local communities before prospecting for natural products. At least two companies, Glaxo Wellcome and Novo Nordisk, have said they will cut back on research on natural products if other countries follow suit.

The bioprospecting working group will also look into the issue of intellectual property rights. Some developing countries, led by Ethiopia, had wanted a firm statement on perceived inconsistencies between the trade-related aspects of intellectual property rights under the World Trade Organization (WTO) rules, and under the biodiversity convention.

The convention calls for intellectual property rights regulations to recognize the contributions of, for example, traditional healers in the development of drugs from natural products. But this is not the case under WTO rules.

Ehsan Masood

Japan's quake strategy urged to switch to long-term forecasting

[TOKYO] Earthquake research in Japan should aim at making 10-year forecasts of locations likely to be struck by major tremors, a group of 160 seismologists has recommended.

The proposal, part of a new research plan drawn up by the Japanese seismologists, represents a break from the previous policy, which has focused on short-term prediction of the location and size of earthquakes likely to occur within a few days.

The new plan is expected to serve as the backbone for the latest earthquake-prediction strategy being drafted by the Geodetic Council, an advisory body to the Ministry of Education, Science, Sports and Culture (Monbusho).

According to Yozo Hamano, leader of the earthquake researchers' group and a professor of Earth dynamics at Tokyo University, the ad hoc group was formed last year in response to a government report on earthquake prediction.

Subcommittees of the Geodetic Council had compiled a report last June which admitted that the short-term prediction of earthquakes — a programme that receives an annual budget of more than ¥20 billion (US\$149 million) — is extremely difficult, and that there is a huge gap between the public's perception of the ability to predict



Tremor terrors: longer forecasts would aim to give more warning of quakes than Kobe had in 1995.

earthquakes and the actual performance of science (see *Nature* 388, 4; 1997).

The new plan, with its long-term strategy, would mark the first change in the 30-year-old programme, which has been focusing mainly on predicting imminent earthquakes in the Tokai region, southwest of Tokyo.

According to Hamano, the new approach emphasizes elucidating a quake plan for the whole Japanese archipelago, through specifying the regions that have crustal

distortions and are likely to experience an earthquake. "Thirty years of research in earthquake prediction has helped collect data for building new research methods," says Hamano. "Recent improvement in observational methods, for example global positioning systems and other space-based techniques, are also encouraging for the new earthquake-prediction programme."

But critics of the prediction programme are sceptical about the scientific basis of the plan, as well as its claim that it would bring changes to the existing programme. The main focus of the programme will see a shift towards long-term forecasts. But research in short-term prediction is still widely supported, and is likely to persist on a smaller scale in the eighth five-year prediction plan, which begins next year.

Robert Geller, a seismologist at Tokyo University, points out that the proposed plan lacks adequate theory or method required for making the reliable forecasts that it claims are possible.

"Accurate earthquake prediction is impossible, and will be for the foreseeable future," says Geller. "What's needed is basic research on the earthquake-source process, rather than a national project aimed at issuing predictions."

Asako Saegusa

France seeks scientific entrepreneurs

[PARIS] A multi-billion-franc package of measures to stimulate innovation and entrepreneurship in France was launched last week by the Socialist government. It includes lifting a ban on public researchers from holding financial interests in companies with which they have research links.

Other measures include FFfr1 billion (US\$167 million) for new national networks of public- and private-sector laboratories in key technologies, and the creation of a FFfr100 million 'seed money' fund to help entrepreneurs to take ideas to the stage where they can seek venture capital.

The government acknowledged what many observers have long been arguing: that France's centralized state planning of technologies is now obsolete, despite its success in the past in areas such as nuclear power, aerospace and transport.

The prime minister, Lionel Jospin, said that this model is "no longer adapted to a global economy in which the market plays a determining role, and where the evolution of knowledge and technology has accelerated".

He said the culture of French science and industry must be changed accordingly, while protecting basic research. The need is to "cultivate the taste for risk and entrepreneurship". French industrial culture is notoriously averse to risk, largely because public-sector researchers enjoy civil-servant status and the engineering élites have secure jobs.

Jospin was speaking at a national conference on innovation in Paris, attended by hundreds of leading researchers, industrialists and businessmen. State intervention has to be "profoundly modified", he said.

There has to be a move away from central planning to a more limited role in supporting education and fundamental research, and in providing a fiscal and legal environment conducive to entrepreneurship. France has to shift from a "logic of subsidies to a logic of incitation", said Jospin.

He announced that direct subsidies to industrial research in large companies will end. Funding will instead be awarded to competitive project proposals that include small companies. Claude Allègre, minister for national education, research and technology, said the need is to shift from an "old-fashioned Colbertism to an enlightened Keynesianism".

The shift in official thinking has been widely welcomed. Daniel Muzyka, professor of entrepreneurship at the prestigious INSEAD business school on the outskirts of Paris, describes the policies as "a very serious step in the right direction".

Many politicians acknowledge that state interventionism has left France with centralized élite *grandes écoles* (see *Nature* 393, 102; 1998), massive public research organizations



Strauss-Kahn: a moving force.

and large state-backed companies. The structure is poorly adapted, in particular, to fostering contacts between scientists, entrepreneurs and small companies.

The new emphasis on fostering contacts marks a significant shift towards considering entrepreneurship and the growth of companies as the key issues, rather than just innovation, the production of ideas and prototypes, says Muzyka.

The glaring gap between France's strong science base and its poor performance in wealth creation was castigated in a recent report to the government by Henri Guillaume, honorary president of the national innovation agency ANVAR (see *Nature* 392, 214; 1998). Picking up this theme at last week's meeting, Allègre deplored the fact that only a handful of the tens of thousands of public researchers transfer to industry every year.

In a move designed to encourage links between public- and private-sector researchers, Allègre and Dominique Strauss-Kahn, the industry and finance minister, announced the creation of a FFfr1 billion fund over three years to create national networks between public and private laboratories in key technology areas.

The government also promised to break a long-standing taboo in France by passing a law before the end of the year to lift the ban on publicly funded researchers holding shares — or stock options — or sitting on the boards of companies with which they have research links.

The move is considered long overdue by scientists. It is aimed at sending a strong political message to government officials, research administrators and scientists that money-making by researchers — long considered a shameful activity in France — is to be actively encouraged as an essential element in wealth creation.

Allègre also suggested that technology transfer and industrial experience should be included in the criteria used to evaluate researchers for promotion and funding. But he failed to give details of how this would work in practice, saying only that measures are under discussion.

Several observers are critical of what they describe as the lack of concrete measures to encourage technology transfer, such as the adoption of strong foresight initiatives or the introduction of a fully fledged postdoctoral system to increase mobility and flexibility.

Strauss-Kahn confirmed a decision to use FFfr600 million of public funds to boost venture capital. The minister said that 5 per cent of Assurance Vie, a popular endowment scheme, will be invested in shares in high-technology companies — a sum that should amount to several billion francs (see *Nature* 392, 856; 1998).

He declared that innovation is his "top priority," arguing that the creation of companies, for example in biotechnology and computing, is now recognized as the engine of economic growth and job creation.

Some critics, such as Marc Giget, head of the Paris-based technology consultancy Euroconsult, argue that the government is failing to abandon interventionism completely. His remedy, shared by several speakers at last week's meeting, is a drastic liberalization of the research and university systems, and a marked withdrawal of the state from industrial policy.

Declan Butler

Report released on INSERM laboratory

[PARIS] An inquiry by the French research ministry into the activities of a laboratory of INSERM, the national biomedical research agency, headed by Bernard Bihain, has taken a new turn. Last week the ministry released a controversial — and previously confidential — report on the laboratory drawn up last year by an independent commission of inquiry (see *Nature* 391, 519 & 825; 1998).

The inquiry, chaired by Pierre Corvol of the Collège de France, received testimony from 24 'whistleblowers' who work or have worked at the INSERM Laboratory of Nutrition, Lipoprotein Metabolism and Atherosclerosis at the University of Rennes. It concludes that the testimonies of seven of these witnesses raise doubts about the

"validity of certain results published or in the process of publication by the director of the laboratory".

A scientific annexe to the report, by John Chapman, a panel member and the director of INSERM's Laboratory of Lipoproteins and Atherogenesis in Paris, claims that Bihain "lacked scientific rigour", and describes allegations that he had "favoured and selected" certain results "to reinforce a hypothesis *a priori*".

Daniel Nahon, director-general of the research ministry, announced last week that four international experts had been asked by the ministry to carry out a new inquiry within three months. He promised that their findings would be made public.

D.B.

G8 leaders pledge to sign Kyoto protocol on climate change

[LONDON] The Group of Eight leaders of the industrialized world last week promised to sign the Kyoto protocol on climate change by the end of 1999; Britain and France have already done so. The leaders also promised at the end of their summit meeting in Birmingham, England, to reduce "significantly" domestic emissions of greenhouse gases.

The second pledge was made following concern from the United Kingdom and other European Union (EU) states that the United States might make most of its reductions by buying up surplus emissions entitlements from countries such as Russia.

EU member states, meanwhile, are still debating how to divide their Kyoto greenhouse-gas reduction target of 8 per cent. A decision is expected at the next meeting of environment ministers at the end of June. Britain and Germany do not want any country to be allowed to increase emissions by more than 10 per cent.

Germany receives fewer visiting research fellows

[MUNICH] Germany's Humboldt Foundation supported more than 2,000 visiting scientists from 90 countries in 1997, according to its annual report published last week. China headed the list with 177 out of the total of 2,212, followed by the United States (155), Russia (131), India (124) and Japan (93).

But the foundation granted only 464 new research fellowships in 1997 compared with 600 in the previous year, said its president, Reimar Lüst, as a consequence of decreases in Germany's budget for foreign cultural and educational policy. He criticized these cuts at a press conference. "If Germany is to keep up its international attractiveness for foreign scientists, it has to become more internationally minded," he said.

Molecular medicine to fore in fight against pain

[MUNICH] Molecular medicine is to play a larger role in the fight against chronic pain and rheumatism in Germany. All six so-called 'leitprojekte' (leading projects) picked as winners of a competition for funding priority in molecular medicine focus on these two research areas.

"Seven million Germans suffer chronic pain, and one half of the older generation suffer from rheumatism," said Germany's federal research minister, Jürgen Rüttgers, when announcing the competition's results last week.

The competition was launched in 1996 as

part of a wider campaign to increase efficiency of research spending (see *Nature* 384, 500; 1996). A DM125 million (US\$70 million) fund has been reserved for the winners over the next five years.

France's young people back careers in science

[PARIS] Scientific research benefits from a positive image among young French people, according to a survey commissioned from the pollster SOFRES by the French government to coincide with a national conference on innovation (see page 203).

Of the 404 people aged between 18 and 30 who were polled, 78 per cent considered a career as a scientist "attractive for a young person", and 91 per cent thought it brought increased social standing and job satisfaction. But one-third thought that a research career was badly paid, and almost two-thirds thought that school education failed to develop a "taste for research and innovation". Almost all those questioned considered that public research should be spared from spending cuts.

Japan to increase spend on CERN collider project

[LONDON] Japan has agreed to contribute a further ¥5 billion (US\$37 million) to the Large Hadron Collider project at the European Laboratory for Particle Physics (CERN) in Geneva, Switzerland. If approved by Japan's parliament, this sum will bring Japan's total contribution to ¥13.85 billion.

Japan was given observer status by CERN three years ago, one month after its first contribution of ¥5 billion, and it provided a further ¥3.85 billion last year (see *Nature* 385, 565; 1997). One hundred and sixty-three Japanese scientists are currently working on CERN projects.

US report gives warning on overuse of antibiotics

[WASHINGTON] A US Institute of Medicine panel reported last week that widespread, inappropriate use of antibiotics has led to the premature emergence of microbial resistance, and that concerted efforts are now required to contain it. Resistance is "newly worrying because it is accumulating and accelerating, while the world's tools for combating it decrease in power and number", the report says.

It recommends better surveillance of drug resistance and more research into antibiotic overuse and misuse, in humans and in agriculture. The report was prepared by the Institute of Medicine's Forum on Emerging Infections, chaired by the Nobel laureate Joshua Lederberg of the Rockefeller University in New York.

Whaling watchdog sinks plan for secret ballots

[LONDON] The International Whaling Commission's (IWC) annual meeting, held in Oman this week, defeated moves by Japan and six Caribbean states to introduce voting by secret ballot. Japan and Norway are the only IWC members openly opposed to a moratorium on whaling.

The Caribbean states argued that secret ballots would allow them to vote "free of fear", but environmentalist groups claim that Japan is lobbying Caribbean countries to support its position, and that a secret ballot would have made this possible. The environmentalists — together with the European Union and the United States — said that secret ballots would have reduced transparency in the IWC process.

Nathanson takes top job at AIDS research office

[WASHINGTON] Neal Nathanson, a virologist and former chair of microbiology at the University of Pennsylvania Medical Center, has been appointed as the new director of the US Office of AIDS Research at the National Institutes of Health (NIH). Nathanson will take over from Jack Whitescarver, who has been acting in the post since William Paul vacated it last November.

NIH director Harold Varmus said that the appointment of Nathanson, who has served on NIH's AIDS Vaccine Research Committee, will "enhance our deep commitment to vaccine research". AIDS activist groups, who support a strong Office of AIDS Research as a counterweight to NIH's institute directors, welcomed the move. "Nathanson is an excellent choice," says Gregg Gonslaves of the Treatment Action Group in New York.

WHO campaign aims to halve malaria deaths



[LONDON] Gro Harlem Brundtland (left), the former prime minister of Norway who is now director-general of the World Health Organization (WHO), has promised to make the fight against malaria a priority of her five-year

term. The United Nations agency's 'Roll Back Malaria' campaign aims to halve deaths from malaria by 2010, and halve them again by 2015. Brundtland's plans were endorsed by last week's Group of Eight meeting of the leaders of the industrialized world.

Britain announced that it would be contributing £60 million (US\$98 million) to the campaign. The United States announced a budget of \$250 million over five years.

Swiss democracy has its advantages

Sir — Swiss democracy clearly differs from both the British parliamentary model and the American system, in which the maximization of individual ends is constitutionally entrenched.

Contrary to your comments about the forthcoming referendum on 7 June on the proposed ban on the production and distribution of transgenic animals, field trials with genetically modified organisms and the patenting of genetically modified animals and plants (*Nature* **392**, 741; 1998), there has indeed been wide public debate in Switzerland.

The potential benefits and risks have been discussed in widely read newspapers. There have also been the demonstrations, advertisements and campaigning that accompany any public vote.

Universities and research and technical institutes have held open days where citizens have come to see at first hand what molecular biology, genetic engineering and biotechnology represent. Municipalities have invited experts — predominantly scientists opposed to the ban — to lecture at community meetings. The 'No' side has been organized and provided with information. We cannot agree, therefore, that "few can have detailed knowledge of the issues at stake".

The government, lobbying groups and international organizations clearly have opinions about the issues under debate. Both sides have presented their views and permitted a public evaluation. As you admit, there is a risk, as is not unknown in political or scientific arenas. But Swiss democracy is one where 'microindicators' at the individual or community level complement macroscopic indicators, such as per capita gross domestic product (GDP).

The passing of the initiative may have a long-term negative effect on these macroindicators. But it may also be that the factors that measure quality of life for many of the population — such as the ability of rural inhabitants to practise traditional living, access to education and distribution of wealth — remain unaffected. There are strong advantages in the way in which the Swiss have, for the past several centuries, managed their democracy.

One argument that can be made against participative democracy is that the public may not be able to assess the effects on macroindicators such as GDP. But the argument in favour of referendums is that national and international institutions or corporations — as well, in this case, as the scientific community — may not be qualified to judge the impact of an alternative on such indicators of quality of

life. In the end, it comes down to a choice of values.

Swiss participative democracy values micro-based quality-of-life indicators as complementary to macroeconomic indices. We do not claim that this choice is better than that of other countries. However, our participative democracy is designed on the basis of this — perhaps uniquely Swiss — preference. Whatever the results of the referendum, "the appropriateness of decision-making by referendum in a highly complex world" is consistent with Swiss priorities. To judge it according to other democratic standards, regardless of how widely they are accepted internationally, is to ignore national autonomy, and all the consequences such choices imply.

John C. Badoux

(President)

David Hunkeler

Laboratory of Polymers and Biomaterials,
Ecole Polytechnique Fédérale de Lausanne,
CH-1015 Lausanne, Switzerland
e-mail: jean-claude.badoux@pres.adm.epfl.ch
hunkeler@igc.dc.epfl.ch

Sir — In the coming vote on a new article in the Swiss constitution on issues of gene technology for non-humans, the result will be decided by the votes of a majority of the population only in combination with the votes of the majority of cantons. If the proposed ban is to take effect, more than half of the citizens and more than half of the 26 cantons would have to accept it. It is rare in Swiss history for such a constitutional referendum to be passed.

Second, the Swiss population is capable of judging complex questions, even in a scientific context. The vote on nuclear energy eight years ago denied a ban on nuclear plants and approved a ten-year moratorium on building new facilities. In the coming vote, the Swiss population has no moderate alternative to approve. Several groups in parliament have suggested alternatives, but industry in particular has lobbied successfully against them. The legislation that the government offered instead is scattered in nine different laws, and is therefore not transparent for people and not easily controlled.

Third, promoters of the ban have raised around SwFr5 million (US\$3.4 million) for their campaign, whereas its opponents have a budget of SwFr35 million. Both sides use posters and slogans and try to appeal to the emotions of citizens. Such campaigns are usual in Switzerland before votes with important implications.

So I do not believe that the Swiss democratic instruments are responsible for

the difficult situation with the vote on genetic engineering. Even if parts of the population do not understand scientific matters, it is not a lack of knowledge or reflection that has created support for the ban. Scientists failed until recently to try to gain public trust. Policy-makers clearly underestimated the danger of not presenting a moderate alternative for the vote or a coherent law. A controversial technology has to be negotiated to get acceptance by the public — not only in Switzerland.

Rosmarie Waldner

Stuhlenstr. 23,
CH 8123 Ebmatingen, Switzerland
e-mail: rosmarie.waldner@tages-anzeiger.ch

Sir — I agree with your advice to the Swiss that passage of their constitutional amendment on aspects of biotechnology would be a serious error. Your further advice about their form of government, however, may well harm the interests of science.

If the Swiss, by referendum, approve these draconian limits, you argue that the blame should be apportioned in part to their system of government. The Swiss version of democracy, though by no means perfect, has been in place during an era when the Swiss have enjoyed levels of personal freedom, security and wealth that would be the envy of most of the world. A careful analysis of the consequences of direct democracy through referendums might show that it is, on balance, undesirable, but your leading article neither attempts nor cites any such analysis. Your position seems to be simpler: if a popular vote would lead to the wrong decision about something as complex as science, it should not be held.

Science is necessarily part of the political system; it is and must be bound by the laws of the countries in which it works. Those opposed to the Swiss biotechnology initiative should work hard to defeat it, not complain about the rules that made it possible. If, without violating fundamental human rights or international agreements, the Swiss choose to forgo much biotechnology, such a choice, though foolish, is theirs to make. By questioning the legitimacy of that choice, you only strengthen an existing view that science is unwilling to accept any outside control, no matter how democratically imposed. I doubt that this helps Swiss opponents of the initiative; I fear it may harm the interests of science throughout the world.

Henry T. Greely

Stanford Law School,
Stanford, California 94305-8610, USA
e-mail: hgreely@stanford.edu

Loopholes exploited in whaling regulation

Sir — When asked what he thought of Western civilization, Mahatma Gandhi is said to have replied: “That would be a good idea”. Simmonds and Stroud¹ ask the scientific community to support their call for a global sanctuary to protect whales from direct killing in all maritime waters; some of us think that, too, would be a good idea, especially as your correspondents appear to wish also to stop what the International Whaling Commission (IWC) quaintly calls “aboriginal subsistence whaling” as well as the much larger scale commercial and “scientific” kinds.

But we have to ask what exactly is this global sanctuary? After all, scientists consider themselves to be rigorous in such matters. Ireland has called for one covering all offshore waters, essentially those beyond national jurisdictions, and as part of a proposed package deal to end the enduring impasse between whaling and non-whaling countries. That I understand, but Simmonds and Stroud don't mean that. If they want a sanctuary to stop all commercial whaling everywhere, then one has already been agreed: it is called an indefinite moratorium, and it was agreed in 1982, coming into force in 1986. The trouble is that under applicable international law there are loopholes that are ruthlessly exploited by Norway and Japan and effectively negate the 1982 decision. And Japan does not even honour the sanctuary declared in the Southern Ocean in 1994.

IWC decisions are also not applicable to countries that do not belong to the organization and, further, the IWC cannot prevent its member states walking out whenever they feel like it.

I wonder under what legal mechanism Simmonds and Stroud would wish to see their “global sanctuary” established and how would they propose that any declaration of it be made binding on all states? By the United Nations, perhaps?

The scientific community, to whom Simmonds' and Stroud's call is addressed, could instead better rest its support for whale conservation and protection on science. The IWC has agreed on the most thoroughly science-based, the most fully tested procedure for regulating use of a living marine resource that has ever been proposed; it is called the Revised Management Procedure (RMP) and it is extraordinarily precautionary in approach². If applied to all whale species and stocks it would, considering their states and the conditions of past exploitation, give zero commercial quotas for nearly all of them,

for many, many years to come. Non-zero quotas would emerge for minke whales, but most of those would be protected if Japan were persuaded to honour the Southern Ocean sanctuary, because that is where 80–90% of the world's minke whales live.

So for the scientific community the rational approach is (1) to secure the protection of Antarctic waters; (2) to apply the RMP elsewhere; (3) to complete the proposed Revised Management Scheme (which includes the RMP) so that there is convincing inspection and international control over all whaling, including — as protection against unlawful trade — a DNA database, to be held by the IWC, and containing information about every whale legally killed.

Sidney J. Holt

Podere il Falco,

06060 Ponticelli, PG, Italy

e-mail: sholt/0002218557@mcimail.com

1. Simmonds, M. & Stroud, C. *Nature* **392**, 541 (1998).

2. IWC The Revised Management Procedure (RMP) for Baleen Whales. Rep. int. Whal. Commn **44**, 145–167 (1994).

We aim to refresh science, not to rebel

Sir — We appreciate your interest in spontaneous meetings organized by young postdocs (see *Nature* **392**, 211; 1998), which demonstrates your interest in the growth and formation of scientists in their early years. You have drawn attention not only to a new initiative but also to some of the problems experienced by postdocs. As organizers of the First European Workshop on Cell Death, however, we fear that the tone of the leading article may open the door to misinterpretation. We should therefore like to clarify our position.

Our meeting is not a ‘rebellion’ but a different type of approach, which we believe is necessary in a rapidly developing field with more than 6,000 publications a year and only one official European meeting. It is not intended to be a competitor of meetings organized by established groups, but a complementary initiative, which we believe fits in the objective of the European Cell Death Organization (ECDO).

We would also like to acknowledge that the European School of Haematology/ECDO awards scholarships to young scientists. Nevertheless, the organization of meetings in cheaper locations with longer discussion time will benefit science by facilitating exchange, confrontation of ideas and collaborations.

We hope that young European scientists will help to ‘refresh’ science rather than ‘rebel’ against it. Our intention is not to create a negative attitude, but to send out a positive signal by showing that it is feasible

to organize an inexpensive workshop, where we will try to create an atmosphere of open and constructive discussion for the improvement of the field.

Vincenzo De Laurenzi

IDI-IRCCS, Biochemistry Laboratory,

University of Rome Tor Vergata,

00133 Rome, Italy

e-mail: delaurenzi@utovrm.it

Michael Hahne

Institut de Biochimie,

University of Lausanne,

Ch. des Boveresses 155,

Epalinges,

CH 1066 Switzerland

e-mail: Michael.Hahne@ib.unil.ch

Henning Walczak

Tumorimmunology Programme,

German Cancer Research Center,

In Neuenheimer Feld 280,

69120 Heidelberg,

Germany

e-mail: h.walczak@dkfz-heidelberg.de

Jan Paul Medema

Immunohematology and Bloodbank,

Leiden University Medical Center,

Albinusdreef 2,

2333AA Leiden,

The Netherlands

e-mail: j.p.medema@

immunohematology.medfac.leidenuniv.nl

BSE coordinator should have been full-time job

Sir — In your article about the bovine spongiform encephalopathy (BSE) inquiry (*Nature* **392**, 532–533; 1998), you comment on my view, expressed in 1991, that “the problem was so big that it needed a coordinator to take hold of the whole thing”. You have, however, missed the essential point I was making, that the problem required a full-time coordinator.

As I expressed in a letter to Keith Meldrum, the Chief Veterinary Officer at that time, part of which you include in your article, I felt it was expecting too much of a group of people (the Spongiform Encephalopathy Advisory Committee headed by David Tyrrell) meeting every two months or so to coordinate the work on BSE.

At no time was I critical of Tyrrell, who I consider did an excellent job. Seven years later, I still hold the view that the appointment of a full-time coordinator, who really knew about spongiform encephalopathies, would have been valuable.

F. Brown

US Department of Agriculture,

Plum Island Animal Disease Center,

PO Box 848,

Greenport, New York 11944, USA

Sound grounds for computing dendrites

Idan Segev

Dendrites are projections that typically originate from the cell body of neurons and are the main site for incoming synaptic inputs. Their function is largely unknown. But there is now clear-cut evidence that, in the auditory brain stem, dendrites enrich the computational power of neurons.

Why is the brain built not from spherical neurons, but from neurons that bear elaborate dendritic trees? These exquisite structures have a different morphology in different parts of the brain (Fig. 1); they are the main receptive region for synaptic inputs from other neurons and (in the periphery) they receive direct input from the sensory system involved. We have long been acquainted with the many faces of dendrites, and with new optical techniques have started to probe into the details of their physiology^{1,2}.

But not much is known about what dendrites actually do. Are they just to enable the connection of synaptic inputs, or simply the result of structural constraints? Are they mere chemical compartments in which very local changes in the efficiency of synapses take place? Or might dendrites also carry out some specific computation that would otherwise be hard, or less efficient, to perform — in short, can dendrites enrich the computational power of neurons? On page 268 of this issue³, Agmon-Snir *et al.* show that they can indeed. Using the rare example of the dendrites of coincidence detector (CD) cells in the auditory brain stem, they show that, together with their synaptic architecture, these dendrites are designed for improving sound localization.

There are several reasons why CD cells provide a unique system for exploring fundamental, yet largely open, questions about the computational abilities of dendrites. First, it is known that these cells detect interaural time differences — that is, the different time that sound takes to reach the two ears, which allows the location of the sound to be identified. Second, the morphology of these cells and their synaptic connectivity are well characterized. They have bipolar dendrites, namely, dendrites that emerge from opposite poles of the cell body (Fig. 2, overleaf); and most importantly, inputs from each ear selectively contact only one dendrite. Third, the electrical properties of these cells and their response to input sound are understood.

In an experimentally based biophysical model, Agmon-Snir *et al.* have succeeded in identifying the principles that underlie the improvement of coincidence detection by dendrites. Moreover, they provide a novel explanation for an interesting experimental finding — that the CD cells sensitive to lower-frequency sounds have longer dendrites than do high-frequency CD cells.

An auditory CD neuron performs its job successfully if it fires maximally when, after compensation for interaural time differences by delays within the nervous system, the inputs it receives from the two ears coincide in time. It should fire less when the inputs arise from only one ear, even if those inputs are doubled. With a linear system, these two possibilities cannot be distin-

guished; what is required is some nonlinear mechanism that dampens the auditory input when it arrives from only one ear but not when it arrives from both ears.

The improvement of coincidence detection (and thus of sound localization) by the CD neurons exploits a fundamental biophysical mechanism — the nonlinear summation, or saturation, of excitatory synaptic inputs and, specifically, the exaggeration of this nonlinear loss when the inputs are spatially clustered on one dendrite rather than distributed between several of them⁴. Theoretical studies show that when synaptic inputs are activated close together — that is, on the same dendrite — the local voltage change in this dendrite is larger, and the reduction in the driving force for the synaptic current is more pronounced, than when the synapses contact different parts of the dendritic tree (that is, two different dendrites). In this second case, more depolarizing current is generated by the synapses and a spike is more likely to be generated in the axon leaving the cell body (Fig. 2), signalling that the sound is coming from a particular location.

All in all, designers of CD neurons who really understand the biophysics of synapses on dendrites will use neurons with bipolar dendrites, rather than spherical neurons. They will then make sure that inputs from each ear terminate on different dendrites.

But those designers also have to avoid over-using the mechanism of nonlinear synaptic addition, because saturation can also lead to loss of performance. When

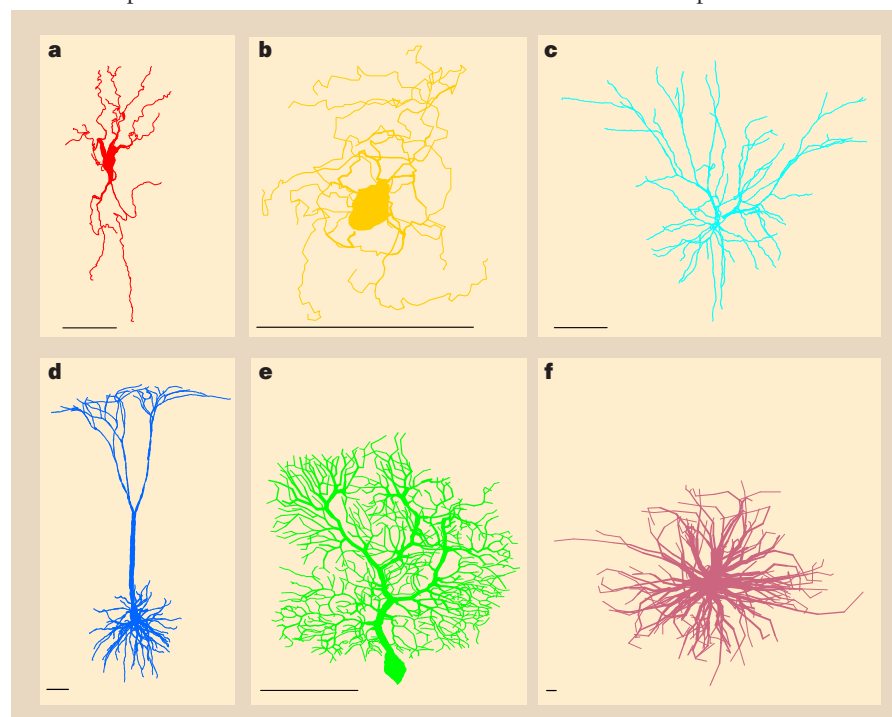


Figure 1 Dendrite variety. Neurons with dendritic trees exist in all sorts of shapes and sizes depending on which region of the brain they come from. Shown here are the dendritic trees of: a, a vagal motoneuron; b, an olivary neuron; c, a layer 2/3 pyramidal cell; d, a layer 5 pyramidal cell; e, a Purkinje cell; and f, an α -motoneuron. Scale bars, 100 μ m.

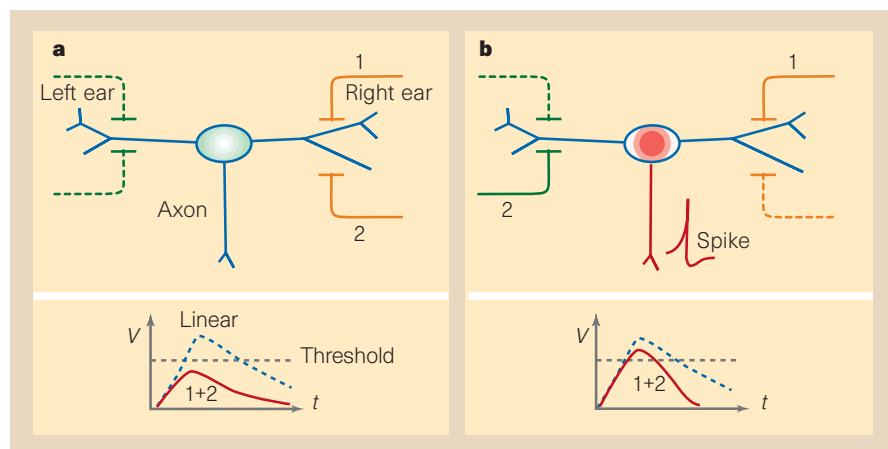


Figure 2 How the bipolar dendrites of coincidence detection neurons, and the mapping of synaptic inputs from each ear to different dendrites, improve sound localization. That improvement is achieved using the nonlinear mechanism (saturation) inherent in the summation of excitatory synaptic inputs, and the exaggeration of this nonlinear loss when the inputs are clustered on one dendrite. **a**, When the input arrives from only one ear, reaching only one dendrite, the consequence of nonlinear summation in that dendrite is that less depolarizing synaptic current is generated. The resulting synaptic potential is too small to generate spike firing in the output axon. **b**, When inputs arrive simultaneously from the two ears and are segregated on different dendrites, nonlinearity in the summation of the synaptic inputs is less significant. The resulting synaptic potential is above the threshold and it gives rise to a spike being fired in the axon.

inputs from one ear are likely to saturate the dendritic potential, a mistimed input from the other ear (to another, unsaturated, dendrite) becomes very effective in depolarizing the axon and may result in an inappropriate firing of the cell. Indeed, with increasing sound frequency, the task of phase-locking between the input sound and the timing of spikes in the auditory nerve is more difficult and the precision of spike timing falls off (that is, the spikes 'jitter'). One way to counteract the possibility of erroneous CD firings is to reduce saturation at the dendrites in high-frequency CD cells, which can be done by making the dendrites shorter. With shorter dendrites, a lower voltage is required at the synaptic site for triggering a spike in the output axon. That is because less current is lost to the other dendrite and more is available to reach the threshold for spike firing; the lower the dendritic voltage, the smaller the saturation at the dendritic site. Obviously, the benefits of nonlinear addition are less for shorter dendrites. So Agmon-Snir and colleagues' model predicts an optimal length for a given frequency, with shorter dendrites for higher frequencies, and experiments⁵ show that this is indeed the case.

What of other types of neuron with dendrites, such as those depicted in Fig. 1 — do they use similar principles to enrich their computational capabilities? The example provided by Agmon-Snir *et al.* is a rare one, because we usually don't know the specific computational function of a given type of neuron. So the modelling studies of the past 30 years need pushing further to continue to provide insights into the possible computational power of neurons with dendrites (for reviews see refs 6–10). The new work³ was

inspired by these studies, and it is likely that the computational module isolated in the auditory neurons for improving coincidence detection might be used in other dendritic neurons with input segregation, such as cortical pyramidal neurons and cerebellar Purkinje cells, perhaps as part of a more complex computational system.

Climate change

The carbon equation

David S. Schimel

Following the signing of the Climate Convention in Rio in 1992, and the subsequent conference in Kyoto late last year, there is a pressing need to find out more about the relationship between anthropogenic emissions of the main greenhouse gas, CO₂, and the resulting atmospheric concentrations. In its reports^{1,2} dealing with 1994 and 1995, the Intergovernmental Panel on Climate Change (IPCC) provided estimates for a wide variety of scenarios, to give policymakers some information on anthropogenic emissions consistent with the aim of stabilizing atmospheric CO₂ at a range of levels (from 350 to 1,000 parts per million by volume).

Uptake or release of CO₂ from the world's oceans and terrestrial ecosystems is central to understanding the relationship between emissions of CO₂ and its atmospheric levels. The IPCC 1994 assessment, however, conducted only a preliminary examination of how altered CO₂, climate and ocean circulation might affect that relationship. Now, in papers on pages 245 and 249 of this issue,

But the findings of Agmon-Snir *et al.* go beyond providing one convincing example of dendritic function, for they also bear on a controversy between two schools of thought. These are the 'individualists', who believe in the functional primacy of the neuronal details (the specific dynamics of synapses, the fine morphology of the dendrites, the exact timing of spikes); and the 'socialists', or 'connectionists', who believe that the characteristics of neuronal society owe little to the properties of its individual components. Agmon-Snir *et al.* have demonstrated the power of the individual, but we eagerly await more studies that tie the variety of dendritic morphologies to their specific function. □

Idan Segev is in the Department of Neurobiology and the Center for Neural Computation, Institute of Life Sciences, The Hebrew University, Jerusalem 91904, Israel.

e-mail: idan@huji.vms.huji.ac.il

1. Stuart, G. J. & Sakmann, B. *Nature* **367**, 69–72 (1994).
2. Yuste, R. & Denk, W. *Nature* **375**, 682–684 (1995).
3. Agmon-Snir, H., Carr, C. E. & Rinzel, J. *Nature* **393**, 268–272 (1998).
4. Rall, W. in *Neural Theory and Modeling* (ed. Reiss, R.) 73–97 (Stanford Univ. Press, 1964).
5. Smith, Z. D. J. & Rubel, E. W. *J. Neurophysiol.* **64**, 465–488 (1979).
6. Segev, I., Rinzel, J. & Shepherd, G. (eds) *The Theoretical Foundation of Dendritic Function: Selected Papers of Wilfrid Rall with Commentaries* (MIT Press, Cambridge, MA, 1995).
7. Mel, B. *Neural Computation* **6**, 1031–1085 (1994).
8. Borst, A. & Egelhaaf, M. *Trends Neurosci.* **17**, 257–263 (1994).
9. Segev, I. in *Handbook of Brain Theory and Neural Networks* (ed. Arbib, M.) 282–289 (MIT Press, Cambridge, MA, 1995).
10. Koch, C. *Biophysics of Computation: Information Processing in Single Neurons* (Oxford Univ. Press, in the press).

Sarmiento *et al.*³ and Cao and Woodward⁴ provide model-based clues to the effects of environmental change on CO₂ fluxes in the oceans and on land. These studies point to substantial differences in the fluxes compared with the estimates used as base cases in the IPCC reports.

Cao and Woodward⁴ suggest that the terrestrial biosphere could act as a significant sink for carbon over the next century because of a large stimulation of photosynthesis by increased CO₂; that sink would, however, be reduced by increased respiratory losses of carbon with warming. The authors present the important result that the terrestrial sink must become saturated — both because photosynthesis follows a saturating function with respect to CO₂, and because, as the rate of increase in photosynthesis slows, plant and microbial respiration must catch up, eventually reducing incremental carbon storage to zero. Although Cao and Woodward's model neglects land-use change and deforestation^{1,2}, and nitrogen deposition⁵, it

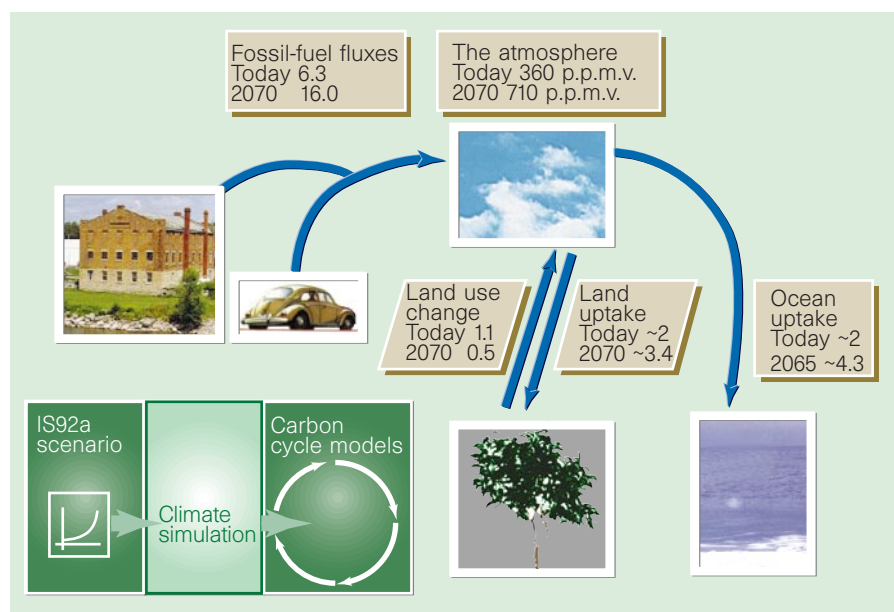


Figure 1 Present-day fluxes of anthropogenic CO₂ compared with estimated fluxes for the year 2070 (or 2065 in the case of Sarmiento *et al.*³). Units are 10^{15} g yr⁻¹ unless stated otherwise; p.p.m.v., parts per million by volume. The estimates for 2070 (2065) are taken from IPCC IS92a figures, or in the case of land uptake⁴ and ocean uptake⁵, are from models of the response of land and oceans to climate change using IS92a as input; IS92a, or IPCC Scenario 92a, gives projections of increasing emissions of CO₂ from use of fossil fuels, assuming moderate growth rates. The inset at bottom left indicates how scenarios of increased anthropogenic emissions of CO₂ feed into climate simulations and then models of the carbon cycle.

demonstrates the maturation of models of plant physiological processes at large scales.

Sarmiento *et al.*³ address the effects of environmental change and marine processes on carbon fluxes in the oceans. The modelled effects occur through changes in salinity, changes in sea-surface temperatures and (via a sensitivity study) changes in the efficiency of biotic uptake on future carbon storage in the oceans. This model links the biota to carbon storage through the availability and utilization of phosphorus as a limiting nutrient, using carbon-to-nutrient (Redfield) ratios. Modelled productivity is controlled by anomalies in the supply of nutrients to the surface ocean, which are caused by changes in circulation or in the efficiency of nutrient use in marine ecosystems. In their somewhat contrived but informative sensitivity analysis, Sarmiento *et al.* show that marine biological changes could have larger effects on ocean carbon storage than alterations in ocean dynamics.

That the fluxes predicted by these models differ from IPCC baseline estimates¹ has serious implications for policy designed to stabilize the concentration of trace gases in the atmosphere. If policies reducing emissions are implemented and CO₂ uptake in the oceans or on land is less than predicted, concentrations will exceed targets. Conversely, if the system is more efficient at sequestering CO₂ than predicted, unnecessary mitigation costs could be incurred. Given the central role of fossil energy, errors in predicting the carbon budget could come at high economic

cost. Both of these studies used only one climate scenario, corresponding roughly to a growth of CO₂ of 1% per year, so the uncertainties in the biogeochemistry they incorporate are amplified by uncertainties in the climate models themselves.

The juxtaposition of these two papers highlights a major difference in approach between marine and terrestrial science. Sarmiento and colleagues' model assumes that productivity is basically linked to nutrient supply and use, and neglects the details of planktonic physiology and population dynamics. Cao and Woodward's model is largely driven by changes to physical resources — light, heat, water and CO₂ — and emphasizes plant physiology and its diversity between life-forms. Yet terrestrial ecosystems are influenced by nutrient supply (as well as internal recycling and adjustment, for instance changes to plant or soil C:N ratios). In particular, further analysis of Cao and Woodward's carbon budget implies that there would be a significant reorganization of the global nitrogen cycle.

Cao and Woodward calculate an atmosphere–land–ecosystem carbon flux of about 1.1×10^{15} g yr⁻¹ in the 1980s and of 3.2×10^{15} g yr⁻¹ in the middle of the twenty-first century. Their estimates go back to 1861, and the overall result is a sink of 300×10^{15} g between then and 2070. The sink is divided about 2:1 between vegetation and soils. This implies a corresponding sink for nitrogen, because organic matter contains, stoichiometrically, both carbon and nitrogen. The sink of nitro-

gen in soils alone, taking a mean of 10:1 for the C:N ratio of soils, implies a minimum sequestration of nitrogen of 30×10^{12} g yr⁻¹ to sustain today's carbon flux and of 100×10^{12} g yr⁻¹ in the future (slightly less if C:N ratios respond to increased CO₂). Storage in vegetation, which is lower in nitrogen than soils, will increase these demands slightly.

These nitrogen fluxes are proportionately quite large relative to today's nitrogen budget (anthropogenic inputs to natural ecosystems today are about $50\text{--}75 \times 10^{12}$ g yr⁻¹), and are unlikely to be sustained in the long term by internal readjustments. Although increased biological nitrogen fixation could make up some of the balance, the implied nitrogen flux to land ecosystems in the years 2050–2070 is more than twice today's estimated rate. Although Cao and Woodward do not directly model nitrogen inputs, their results imply that there will be a significant demand on nitrogen to sustain extra CO₂-stimulated photosynthesis. Some terrestrial ecosystems may be limited by phosphorus, but Alan Townsend of the University of Colorado (personal communication) has calculated an upper bound for carbon storage in tropical forests from phosphorus deposition of no more than 0.1×10^{15} g yr⁻¹, or roughly a tenth of the Cao and Woodward flux — again, this is a large discrepancy to be accounted for by internal readjustments.

Study of the interaction of the biogeochemical cycles⁶ has been eclipsed by the stunning increases in our understanding of terrestrial photosynthesis, but results such as Cao and Woodward's imply a need for a renewed research emphasis on the global cycles of nitrogen, phosphorus and other nutrients to better predict what will happen to levels of atmospheric CO₂. Likewise, the sensitivity of Sarmiento and colleagues' model to nutrient use implies a need for substantially increased biological detail, along the lines of Cao and Woodward's approach, in large-scale ocean models. Taken together, these two studies bring a renewed sense of urgency to research on the future behaviour of the Earth System, especially the coupling of anthropogenic greenhouse gases to the climate and carbon systems. □

David S. Schimel is in the Max-Planck-Institut für Biogeochemie, Sophienstraße 10, 07743 Jena, Germany, and the National Center for Atmospheric Research, PO Box 3000, Boulder, Colorado 80307-3000, USA.

e-mail: schimel@cgd.ucar.edu

- Schimel, D. S. *et al.* in *Climate Change* 1994 (eds Houghton, J. T. *et al.*) 35–72 (Cambridge Univ. Press, 1995).
- Schimel, D. S. *et al.* in *Climate Change* 1995 (eds Houghton, J. T. *et al.*) 65–131 (Cambridge Univ. Press, 1996).
- Sarmiento, J. L., Hughes, T. M. C., Stouffer, R. J. & Manabe, S. *Nature* **393**, 245–249 (1998).
- Cao, M. & Woodward, F. I. *Nature* **393**, 249–252 (1998).
- Holland, E. A. *et al.* *J. Geophys. Res.* **102**, 15849–15866 (1997).
- Bolin, B. & Cook, R. B. (eds) *The Major Biogeochemical Cycles and Their Interactions* (Wiley, New York, 1983).

Molecular neuroendocrinology

Working backwards to find answers

Jean-Claude Meunier

Many of our vital needs — such as those for food and drink, sex, sleep, flight or fighting — depend heavily on the hormone balance, which is tuned by both internal and external stimuli. For example, secretion of hormones by the anterior pituitary gland is controlled largely by the hypothalamus, a region of the brain that lies just above the gland. From the pioneering work of the Guillemain and Schally groups in the 1960s, hypothalamic neurons are known to make and release peptide factors that stimulate or inhibit the secretion of a particular hormone by the specific set of cells that produces it in the pituitary. On page 272 of this issue, Hinuma *et al.*¹ report the isolation of such 'hypothalamic releasing factors' — neuropeptides from rat brain that promote the release of prolactin, the anterior pituitary hormone that is responsible for lactation.

Many factors, including neurotransmitters (such as serotonin) and hormones (thyrotropin-releasing hormone and vasoactive intestinal peptide, for example), are known to stimulate prolactin release². But the newly identified peptides may be the real thing, because they seem to affect the release of no other pituitary hormone. The discovery is interesting not only because it identifies a new and specific level of hormonal regulation that could be developed for therapeutic purposes, but also because it constitutes one of the few examples of successful 'reverse pharmacology', whereby bioactive substances are identified based on a knowledge of the target (receptor), and not the opposite as was the case until recently^{3,4}.

Hypothalamic releasing factors act by stimulating specific receptors in the pituitary cell membrane. These receptors couple with a transducer guanine-nucleotide-binding (G) protein to amplify the intracellular production of a second messenger such as cyclic AMP and promote secretion of the hormone. To identify the ligands for these receptors, Hinuma and colleagues started virtually from scratch (Fig. 1) — they had no clues like those used to identify nociceptin⁵, the receptor for which shares considerable sequence similarity with the opioid receptors. They first cloned a G-protein-coupled receptor (GPCR) of unknown ligand (hence the term 'orphan receptor') called hGR3. This receptor showed the highest, yet moderate (around 30%), sequence identity with known peptide-activated GPCRs, making it difficult to predict which post-receptor biochemical event(s) its natural ligand could

trigger. But, knowing that the orphan-receptor messenger RNA was most abundant in the pituitary, the authors set out to screen transfected cells that expressed the receptor for an easily quantifiable response to extracts from bovine hypothalamus. They found that, in response to the extracts, the transfected cells promoted the release of arachidonic acid, a cellular lipid metabolite that had been thought for some time to mediate the release of prolactin².

Using this functional assay in combination with chromatographic fractionation of the extracts, Hinuma *et al.* tracked the active components down to two peptides of 29 and 19 amino acids in length, with overlapping sequences. As is the norm for small peptide hormones, the two peptides are synthesized as part of a larger precursor. The authors cloned the complementary DNA for this, and found that it encodes slightly longer peptides of 31 and 20 amino acids, with carboxy-terminal amidation motifs (a glycine residue followed by a proteolytic cleavage signal; in this case a triplet of arginyl residues), as is also often seen in peptidic hormones.

Provided they have this carboxy-terminal amidation, the synthetic peptides — which the authors refer to as PrRP31 and PrRP20 —

show potent prolactin-releasing activity in a variety of physiologically relevant situations. The peptides stimulate the release of prolactin from cultured pituitary tumour cells, healthy lactotrophs (the pituitary cells that make and secrete the hormone) in primary culture, and, *in vivo*, from the rat anterior pituitary in a perfusion assay. But are the two peptides both physiologically significant? The shorter (PrRP20) peptide may be naturally derived from the longer one by complete processing of the precursor at a single arginyl residue. To resolve this, we could test whether the two peptides are produced in distinct sets of neurons, or whether they are differentially secreted in distinct physiological situations such as pregnancy and lactation.

Clearly, the discovery of Hinuma *et al.* opens up new perspectives in the field of neuroendocrinology and, as is always tempting to speculate, in neurophysiopathology. According to the neuroendocrinologist Jacques Epelbaum (personal communication): "Prolactin displays hundreds of functions during the course of evolution. However, one can now search for and consider the use of antagonists of the hGR3 receptor to treat cases of hyperprolactinaemia, especially those resistant to bromocryptine. One could also seriously consider the possibility that mutations turning the hGR3 receptor to permanently active, even in the absence of releasing factor, may be involved in the physiopathology of pituitary adenomas". In much the same way, activating mutations of other GPCRs (such as the thyrotropin and luteotropin receptors) cause hyperthyroidism⁶ and precocious puberty⁷, respectively. ►

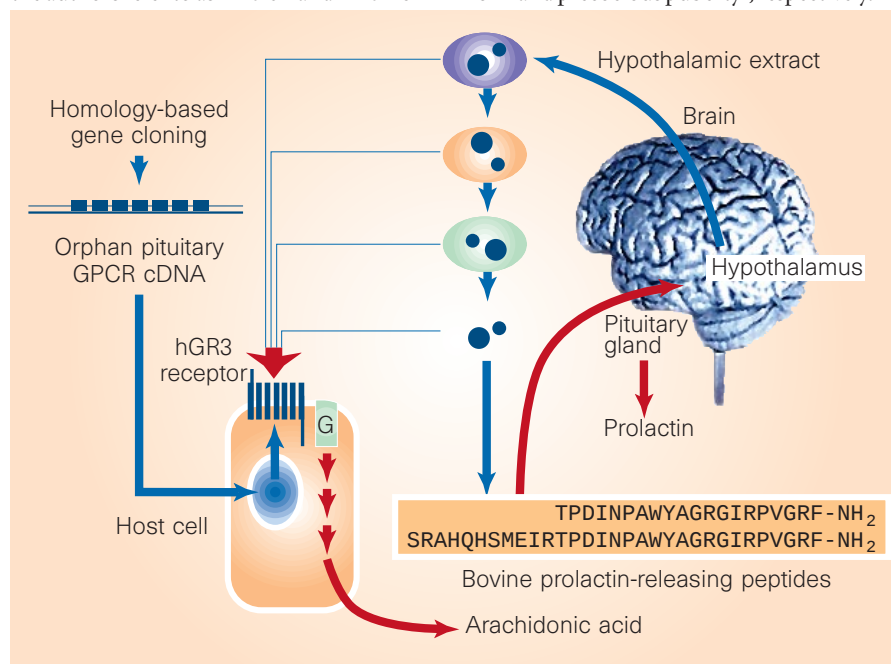


Figure 1 The strategy used by Hinuma *et al.*¹ to isolate and identify the prolactin-releasing peptides. The components which, in the hypothalamic extracts, turned out to be the active ones (that is, promote synthesis of arachidonic acid in recombinant host cells) are depicted as small and larger black circles. The sequences are those deduced from the primary structure of the precursor, the complementary DNA for which the authors have also cloned.



100 YEARS AGO

We learn from the *Lancet* that the use of Röntgen rays as a means of certifying the existence of death was demonstrated at a recent meeting of the Biological Society of Paris. M. Bougarde showed three photographs of the thorax, two of them from living persons and the third from a corpse, all taken by the X-rays. In the two first the different thoracic organs and the walls of the thorax itself exhibited a hazy outline, so that their limits could not be exactly made out. This, of course, was owing to the natural movements of the parts, the pulsations of the heart and the great vessels, and the movements of the diaphragm. Even when the subjects held their breath so as to minimise movement as much as possible the outlines were still hazy. ... In the radiograph of the corpse, however, the appearance was quite different, for all the organs had sharp and well-defined edges.

From *Nature* 19 May 1898.

50 YEARS AGO

Although lip-service is commonly paid to the value of research, it is seldom that its direct benefits are so attractively displayed to the general public as has been done in the "Darkness into Daylight" Exhibition at the Science Museum, London, S.W.7. ... We are all of us aware of the progress and achievements made during the past century. But how many realize that since only 1921 the efficiency of the ordinary 60-watt electric lamp has increased by 55 per cent while the cost has decreased by 75 per cent. ... We can see the dawn of electric lighting with Swan's invention of 1878, its temporary eclipse by the Welsbach incandescent gas mantle and then the final triumph of the electric filament lamp. ... As is perhaps inevitable, the accent of the Exhibition is on fluorescent lighting, and as one passes through this section one is forced to wonder whether the life of the filament lamp is doomed and whether it will eventually be replaced, like the oil lamp and the gas lamp before, by the low-pressure fluorescent tubes now being so widely used. Hitherto, these fluorescent lamps have only been available in lengths of 4 and 5 ft., but the recent announcement that 2-ft. tubes of 20- and 40-watt ratings will soon be introduced suggests a much wider application for domestic purposes.

From *Nature* 22 May 1948.

Reverse pharmacology based on orphan GPCRs is a unique opportunity for drug discovery, especially given that most known therapeutic agents have GPCRs as their main target. The approach has proved itself on four occasions — for nociceptin⁵ (orphanin FQ)⁸, lymnokinin⁹, the orexins¹⁰ and now the prolactin-releasing peptide(s). Considering that over 100 orphan GPCR sequences have already been identified, and that when the Human Genome Project is completed, hundreds, if not thousands, more will probably have been found⁴, no doubt this approach will soon provide many new leader compounds for the pharmaceutical industry to convert into clinically effective agents. □

Jean-Claude Meunier is at the Institut de Pharmacologie et de Biologie Structurale, Centre National de la Recherche Scientifique, 205 route de Narbonne, 31077 Toulouse Cedex 4, France.
e-mail: jcm@ipbs.fr

1. Hinuma, S. *et al.* *Nature* **393**, 272–276 (1998).
2. Lamberts, S. W. & Macleod, R. M. *Physiol. Rev.* **70**, 279–318 (1990).
3. Mills, A. & Duggan, M. J. *Trends Pharmacol. Sci.* **14**, 394–396 (1994).
4. Stadel, J. M. *et al.* *Trends Pharmacol. Sci.* **18**, 430–437 (1997).
5. Meunier, J.-C. *et al.* *Nature* **377**, 532–535 (1995).
6. Parma, J. *et al.* *Nature* **365**, 649–651 (1993).
7. Schenker, A. *et al.* *Nature* **365**, 652–654 (1993).
8. Reinscheid, R. K. *et al.* *Science* **270**, 792–794 (1995).
9. Cox, K. J. A. *et al.* *J. Neurosci.* **17**, 1197–1205 (1997).
10. Sakurai, T. *et al.* *Cell* **92**, 573–585 (1998).

Thermodynamics

Size is everything

R. Stephen Berry

We are familiar with how bulk matter melts and freezes, and know that a pure substance has a sharp, well-defined melting temperature at a given pressure. But in clusters of tens, hundreds or thousands of atoms, very different properties appear. The report by Schmidt *et al.* on page 238 of this issue¹ is the first quantitative demonstration of how the melting temperature of clusters depends sensitively on their size. The temperature varies in what appears to be a highly irregular pattern, a pattern not yet interpreted, and the challenge is now to discover the physical basis for this variation.

The very concept of 'a' melting temperature disappears for such atomic clusters; instead, they have 'melting ranges' of temperature, in which solid and liquid coexist. An earlier publication by this group² showed how the fragmentation pattern of clusters could reveal the relation between their internal energy and their temperature, and hence their heat capacity. This in turn demonstrated experimentally the long-predicted finite range of temperatures within which solid and liquid clusters may coexist^{3,4}. The authors identified the temperature of the maximum heat capacity as the melting point, because this is approximately the temperature at which an ensemble of the clusters is half solid, half liquid.

A size dependence of the melting of sodium clusters was suggested four years ago⁵. Schmidt *et al.* now show that changing the size of sodium clusters by just a few atoms can change their melting temperature by tens of per cent. Why?

The temperature range for coexistence of solid and liquid clusters depends on atomic binding energy — the ease with which atoms can be set free from their lattice sites, to move diffusively or to vibrate. And the binding energies of small and medium-sized clusters do not increase monotonically with cluster

size, because they depend on two 'shell effects'. One is the effect of geometric shells: regular polyhedra, especially icosahedra and decahedra, are exceptionally stable. The other is the effect of electronic shells, in which the exceptionally stable structures correspond to complete electron shells analogous to those responsible for the stability and inertness of the rare gases. These effects compete to determine maxima and minima in the binding energy of clusters as functions of their size⁶. So could it be that shell effects are causing the observed melting-temperature variation? Schmidt *et al.* consider this possibility, but conclude that neither shell effect is strong enough to account for the variations they measured in melting temperature.

What of the entropy change that accompanies the change of phase? At the point of a phase change in bulk matter, the internal energy change from solid to liquid, ΔU , must exactly balance the contribution of entropy to the total energy change, $T\Delta S$ (where T is the temperature, S the entropy). Only at temperatures and pressures where these two are equal can solid and liquid coexist in equilibrium — this is the relation that yields the coexistence curve of the solid-liquid phase diagram. At any temperature and pressure off the coexistence curve, the equilibrium ratio of solid to liquid *bulk* samples in an ensemble is so overwhelmingly favourable to one phase or the other that, if a system is at equilibrium, the unfavoured phase isn't seen. Only by using a trick such as supercooling a liquid can we reveal the local stability of the unfavoured phase.

In clusters, the equality of ΔU and $T\Delta S$ ensures that the solid and liquid forms of the cluster are equally likely to be found. But the equilibrium ratio of liquid to solid clusters need not be unity for the two phase-like forms to coexist in observable quantities.

The changing slope of the internal energy U with temperature, over many tens of degrees Kelvin, displays this coexistence quantitatively for the first time in an experiment. The key to interpreting how these ranges, as well as the points of maximum slope of $U(T)$, vary with size, will come with understanding just what kinds of motion the atoms of liquid-like clusters can undergo.

Another challenge lies ahead. The theory of phase changes predicts that there are sharp limits to the band of coexistence, at any fixed pressure^{3,4}. There should be a lower temperature limit, below which the liquid has not even local stability, and an upper limit, above which the solid has no local stability. Determining whether these sharp bounds exist will be difficult, requiring careful scrutiny of how the internal energy and heat capacity change as the conditions move into the coexistence region. It may well be that clusters of some sizes will show sharp thermodynamic changes and others will show more gradual changes. The clusters with the highest melting temperatures and highest coexistence temperatures will probably be those with the sharpest changes in heat capacity, because these will be the clusters for which the onset of melting is the most abrupt, in terms of the rate at which the

increase in the entropic contribution makes the liquid become a locally stable form to compete with the lower-energy solid.

It is exciting to see experimental studies of the phase changes of clusters go beyond the demonstration that liquid and solid forms exist. Such investigations may do even more than reveal the fundamental character of phase changes. They open the possibility of exploring stable, phase-like forms of clusters — some like conventional solids, some floppy, some possibly unlike any normal forms of bulk solids — that cannot exist as equilibrium forms of bulk matter, and thence to the possibility of preparing and using nanoscale materials with structures unique to small systems, structures with properties that we select and design. □

R. Stephen Berry is in the Department of Chemistry, University of Chicago, 5735 South Ellis Avenue, Chicago, Illinois 60637, USA.

e-mail: berry@rainbow.uchicago.edu

1. Schmidt, M., Kusche, R., von Issendorff, B. & Haberland, H. *Nature* **393**, 238–240 (1998).
2. Schmidt, M., Kusche, R., Kronmüller, W., von Issendorff, B. & Haberland, H. *Phys. Rev. Lett.* **79**, 99–102 (1997).
3. Berry, R. S. *Microsc. Thermophys. Eng.* **1**, 1–18 (1997).
4. Berry, R. S. in *Clusters of Atoms and Molecules* (ed. Haberland, H.) 187–204 (Springer, Berlin, 1994).
5. Martin, T. P., Näher, U., Schaber, H. & Zimmerman, U. *J. Chem. Phys.* **100**, 2322–2324 (1994).
6. Martin, T. P. *Phys. Rep.* **273**, 199–232 (1996).

Behavioural ecology

Cuckoos beg the answer

Rory Howlett

Bawling babies are the bane of many long-suffering parents. But consider the lot of a pair of birds that spends weeks building a nest, and preparing to raise a brood of their own, only to be hoodwinked into caring instead for a gluttonous interloper. That is the fate of victims of the common cuckoo, *Cuculus canorus*. Cuckoos are brood parasites: rather than building nests of their own, they lay their eggs in the nests of other species, relying on the unwitting foster parents to incubate the egg and feed the hatchling until it is fully fledged. But why do the hosts put up with it? Therein lies a sinister and unlikely tale, as told by Nicholas Davies and colleagues in *Proceedings of the Royal Society*¹.

Davies *et al.* are interested in the battle of wits between the cuckoo and one of its host species, the reed warbler *Acrocephalus scirpaceus*. They studied a breeding population of about 300 reed warbler pairs in and around Wicken Fen near Cambridge in England. Reed warblers are vulnerable to cuckoos, and in some years at Wicken nearly a quarter of their nests are parasitized, although in other years only 1–2% of nests are affected. Female cuckoos surreptitiously lay a single egg in an unguarded nest. Adult reed warblers reject eggs that are

unlike their own, and their powers of discrimination have been honed by natural selection. But in this coevolutionary arms race, the cuckoo has also been selected to lay eggs which beautifully mimic those of its

hosts, and which are often accepted and incubated².

Paradoxically, cuckoo chicks are apparently not actively rejected, although they differ in appearance from the chicks of the host species. Not only does the impostor escape detection, it ejects the hosts' own eggs and hatchlings from the nest, with the foster parents often looking on helplessly. Why hosts tolerate these natural-born killers is perplexing, but theory has suggested that learning to recognize cuckoo chicks may be maladaptive for hosts³. Thus, if first-time breeders misimprint on a parasitic chick, thereby failing to recognize future offspring as their own, the costs of nestling recognition could outweigh the benefits in terms of lifetime reproductive success.

Whatever the case, cuckoo nestlings are more than tolerated. Although the foster parents are presented with just a single mouth gape rather than a nest full of gapes of their own young, they strive to provide the ravenous cuckoo chick with a copious supply of caterpillars and insects, leading to rapid growth. Grotesquely, the reed warbler ends up being dwarfed by the chick it is feeding (Fig. 1).

To find out how a cuckoo chick stimulates its foster parents to provide so much food, Davies *et al.* used some neat tricks. To test whether large size is sufficient to stimulate high provisioning rates, they temporarily replaced reed warbler broods with a single blackbird (*Turdus merula*) or song thrush (*T. philomelos*) chick. The surrogate chicks were accepted but were fed less food than cuckoo chicks of similar body mass. So size alone is not the key.

Perhaps cuckoo chicks are just expert beggars. The cuckoo chick has an unusual begging call, which Davies *et al.* describe as a



Figure 1 Fed up — cuckoo chick with a reed warbler surrogate parent.

continuous and rapid 'si, si, si, si ...' and which is quite unlike the begging call of the reed warbler chick. But it closely resembles that made by a whole brood of reed warbler chicks. Sonograms from recordings of chicks induced to beg under controlled laboratory conditions confirmed that the begging call rate of a single hungry cuckoo chick matches the combined calling rate of an average reed warbler brood. So vocal trickery seems a likely explanation for the accelerated provision rate achieved by the cuckoo chick. The real clincher was a field experiment, however, one in which Davies *et al.* used loudspeakers to broadcast cuckoo begging calls or the calls of whole reed warbler broods to nests containing just single blackbird or song thrush chicks. This simple but elegant manipulation causes the foster parents to increase their

provisioning rate to that for a cuckoo chick.

These results provide a beautiful example of sensory exploitation. No doubt future work will explain why the foster parents do not catch on to the fact that they are being taken for a ride. Perhaps it is simply that they are handicapped in the evolutionary arms race with the cuckoo. After all, as Richard Dawkins has observed, every cuckoo is descended from a continuous line of cuckoos that have successfully parasitized hosts, while not every pair of reed warblers falls victim to a cuckoo. □

Rory Howlett is Deputy Biological Sciences Editor of Nature.

e-mail: r.howlett@nature.com

1. Davies, N. B., Kilner, R. M. & Noble, D. G. *Proc. R. Soc. Lond. B* **265**, 673–678 (1998).
2. Brooke, N. de L. & Davies, N. B. *Nature* **335**, 630–632 (1989).
3. Lotem, A. *Nature* **362**, 743–745 (1993).

Neutron stars

A star powered by magnetism

S. R. Kulkarni and Christopher Thompson

Ever since their discovery three decades ago, neutron stars have fascinated astronomers. They have extreme properties in almost all respects: they are the densest objects in the Universe, the most rapidly rotating, and the best astronomical laboratory for condensed-matter physics. To this, Kouveliotou *et al.* (page 235 of this issue¹) add direct evidence of magnetic fields strong enough to perturb the very structure of the vacuum.

First a bit of background. Pulsars — pulsating radio sources — are the most common manifestation of neutron stars. When they are born in supernova explosions, pulsars are believed to be rotating rapidly, with periods of tens of milliseconds, and to have magnetic fields of about 10^{12-13} gauss (10^{8-9} tesla). The source of energy for the emissions

from a pulsar is its rotation: the rapid rotation and the magnetic field combine to generate particles moving at nearly the speed of light, which in turn emit the observed very bright radio waves and high-energy photons. Neutron stars with weaker magnetic fields, 10^9 gauss or even lower, manifest themselves as millisecond pulsars and as accreting X-ray binaries.

There is, however, no reason why a neutron star could not be much more strongly magnetized. A 10^{13} -gauss magnetic field carries no more than a millionth to a billionth of the energy of convection and differential rotation in a newly formed neutron star, and rapid rotation in the millisecond range should lead to an effective large-scale dynamo². More generally, tensile strength does not limit the flux density that a degener-

ate star can support, as is demonstrated by the existence of white dwarfs with magnetic fields approaching 10^9 gauss. But do any neutron stars that start with field strengths well in excess of 10^{13} gauss actually exist?

The story of such 'magnetars' begins on 5 March 1979, when a very bright γ -ray burst was seen towards a bright supernova remnant (SNR) N49 in the Large Magellanic Cloud³. This burst was followed by a 'ringing' — a decaying oscillation of softer γ -ray emission, with a period close to eight seconds. Fainter and much shorter bursts were detected over the next four years. There was considerable controversy about the origin of the source, including whether it was next door in our Galaxy or in the Large Magellanic Cloud. Over the next ten years, it became clear that 5 March 1979 was one of a small group of objects that undergo recurring bursts in the soft γ -ray band, leading to the identification of the class of 'soft gamma-ray repeaters' (SGRs)⁴.

Observations have now established that SGRs are young neutron stars still embedded in their SNR. A key step was the proposed association⁵ of SGR1806–20 with the SNR G10.0–0.3. Rapidly and by luck, an outburst of SGR1806–20 was imaged with the ASCA telescope⁶ and shown to coincide with the strongly peaked radio emission near the centre of G10.0–0.3. This association led to reinvestigation of the 5 March 1979 source, now renamed SGR0525–66, and the discovery⁷ of an X-ray source which was proposed to be its quiescent counterpart. The third member of the SGR class, 1900+14, may also be associated with a supernova remnant^{8,9}.

The apparent diversity of the three SGRs has continued to bewilder astronomers, and raised doubts about their unity as a class. For example, there is no optical counterpart to SGR0525–66, but near SGR1806–20, coinciding with the radio peak of G10.0–0.3 (see Fig. 1), is a luminous blue variable, one of the brightest stars in the Galaxy¹⁰.

In parallel with these primarily observational developments, it was suggested^{2,11–13} that the SGRs are neutron stars with magnetic fields in excess of 10^{14} gauss. The prime distinction between a pulsar and such a magnetar is that the X-ray and particle emissions from the magnetar are powered not by rotation but by the decaying magnetic field. This involves both internal heating and seismic activity that shakes the magnetosphere and accelerates particles. This gradual release of energy is punctuated by intense outbursts that are most plausibly triggered by a sudden fracture of the neutron star's rigid crust.

If SGR0525–66 is indeed an isolated neutron star left over from the explosion that created N49, and it has spun down to an eight-second period in the $\sim 10,000$ years since that event, one can deduce that its field strength is $\sim 6 \times 10^{14}$ gauss. Fields of this strength are also implied by the extreme peak

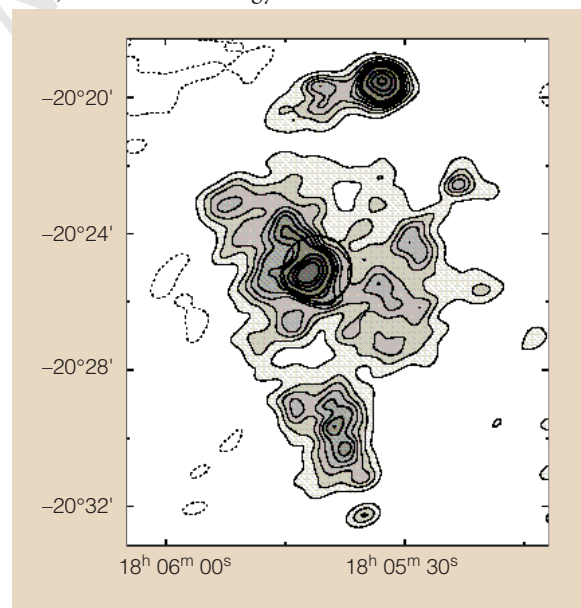


Figure 1 Radio image of the supernova remnant G10.0–0.3 at 1.4 GHz. The circle marks the location of the steady X-ray counterpart of the soft gamma-ray repeater 1806–20. Kouveliotou *et al.*¹ report the detection of 8-second pulsations from this object. A luminous blue variable star coincides with the peak of the radio emission, as discussed in the text. The right ascension and declination are in the B1950 system. (From ref. 18.)

luminosity of the 5 March burst, which released energy at a million times the rate of the Crab pulsar; and by the enormous energy of $\sim 3 \times 10^{44}$ erg that was confined close to the neutron star over the 200-second duration of the burst¹². But these arguments, although compelling and consistent, are indirect.

By contrast, the evidence now presented by Kouveliotou *et al.*¹ is direct. Using the Rossi X-ray Timing Explorer, they have discovered 7.47-second pulsations in the X-ray counterpart⁶ of SGR1806–20. The authors go further in providing two independent measurements of an increasing period, both of which imply a magnetic dipole field somewhat stronger than 10^{14} gauss. They also make cogent arguments against accretion as an energy source for the quiescent X-ray emission, leaving magnetic-field decay as the most likely alternative. The rotational energy of the neutron star with its present period is far too small to power the X-ray and particle emissions.

The important caveat here, raised by the authors, is that if the star seen near SGR1806–20 is its binary companion, then the measured period increase could be modified by the gravitational acceleration of the companion. In addition, a neutron star of this long period, if as hot internally as its quiescent X-ray emissions imply, will undergo sudden decreases in period¹³ (glitches) by perhaps one part in 10,000.

The two SGR sources 0525–66 and 1806–20 are not the end of the story. As we go to press, a possible discovery of 89-second pulsations from SGR1900+14

has been announced¹⁴. The long period may seem puzzling, but it fits in very well with our theoretical expectations. Unlike pulsars, magnetars should have two distinct phases of spindown: the standard pulsar mechanism and a particle-aided spindown. The spin period increases exponentially during the latter phase. The long period of SGR1900+14 is therefore consistent with its low activity — it need only be older than the other two SGRs. Another wrinkle is the identification^{15,16} of a nearby candidate magnetar, RX J0720.4–3125, which is consistent with the inference that magnetars may constitute a sizeable fraction of neutron stars, about 10%.

What does the future hold? First, we must measure changes in period for SGR0525–66 and RX J0720.4–3125. Another important target is the 12-second X-ray pulsar 1E 1841–0.45 in the very young supernova remnant¹⁷ Kes73, one of several anomalous, low-luminosity X-ray pulsars that bear some resemblance to SGR0525–66 in quiescence.

More generally, observations of neutron stars that are slowly rotating but have warm interiors should provide valuable new information on the physics of their interiors, in particular the superfluid component in their crusts. Measurement of low-frequency noise in their spindown is a promising discriminant between accreting and non-accreting behaviour. In the radio regime, observations of long-period, isolated neutron stars will also provide important constraints on emission models.

As the evidence for magnetars grows, it

becomes more plausible that our failure to detect radio pulsars with magnetic fields greater than 2×10^{13} gauss is because, at such high fields, quantum electrodynamic effects help to damp the radio emission. Detailed X-ray spectra obtained by the AXAF satellite will probe radiative transport in strong magnetic fields, and allow neutron stars to expand their roles as laboratories for physics in extreme conditions. □

S. R. Kulkarni is in the Division of Physics, Mathematics and Astronomy, Caltech, Pasadena, California 91125, USA.

e-mail: srk@astro.caltech.edu

Christopher Thompson is in the Department of Physics and Astronomy, University of North Carolina, Chapel Hill, North Carolina 27599-3255, USA.

e-mail: thompson@physics.unc.edu

1. Kouveliotou, C. *et al.* *Nature* **393**, 235–237 (1998).
2. Duncan, R. C. & Thompson, C. *Astrophys. J.* **392**, L9–L13 (1992).
3. Mazets, E. P. *et al.* *Nature* **282**, 587–589 (1979).
4. Atteia, J.-L. *et al.* *Astrophys. J.* **320**, L105–L110 (1987).
5. Kulkarni, S. R. & Frail, D. A. *Nature* **365**, 33–35 (1993).
6. Murakami, T. *et al.* *Nature* **368**, 127–129 (1994).
7. Rothschild, R. E., Kulkarni, S. R. & Lingenfelter, R. E. *Nature* **368**, 432–434 (1994).
8. Vasisht, G. *et al.* *Astrophys. J.* **431**, L35–L38 (1994).
9. Hurley, K. *et al.* *Astrophys. J.* **463**, L13–L16 (1996).
10. Van Kerkwijk, M. H. *et al.* *Astrophys. J.* **444**, L33–L35 (1995).
11. Paczynski, B. *Acta Astron.* **42**, 145–153 (1992).
12. Thompson, C. & Duncan, R. C. *Mon. Not. R. Astron. Soc.* **275**, 255–300 (1995).
13. Thompson, C. & Duncan, R. C. *Astrophys. J.* **473**, 322–342 (1996).
14. Marsden, D. & Rothschild, R. *IAU Circ. No.* 6882 (1998).
15. Motch, C. & Haberl, F. preprint astro-ph/9802301 on xxx.lanl.gov
16. Kulkarni, S. R. & van Kerkwijk, M. H. preprint astro-ph/9803024 on xxx.lanl.gov
17. Vasisht, G. & Gotthelf, E. V. *Astrophys. J.* **486**, L129–L132 (1997).
18. Kulkarni, S. R., Frail, D. A., Kassim, N. E., Murakami, T. & Vasisht, G. *Nature* **368**, 129–131 (1994).

Neurobiology

Knowing where you're going

In terms of mental processes, how do you get from A to B? This question of navigation has been tackled by Eleanor A. Maguire and colleagues, who report their latest findings in the 8 May issue of *Science* (280, 921–924; 1998). Using positron emission tomography they have mapped brain activity as people find their way through a familiar, yet complex, virtual-reality town.

Subjects were allowed to get their bearings in the virtual town (pictured), and they were then asked to head directly towards a goal. Increased blood flow was found in the right hippocampus and inferior parietal cortex of people who successfully completed this task, indicating that these two regions allow navigation to an unseen goal.

Activity was also seen in these areas of the brain when some of the routes to the goal were blocked, forcing the subjects to make a detour. In this case, the left frontal cortex was activated as well, and the authors infer that this region is involved in



planning and making decisions. They also studied average speed, measured in virtual metres per second, and found that it correlates with activity in the right caudate nucleus. So this region probably helps people to get where they're going quickly.

What Maguire and colleagues now have is a map of the areas in the brain that

support navigation. These findings agree with results from monkeys and rats, and also with studies of London taxi drivers asked to recall complex routes around the city — all of which goes to show that some experiments on rodents really do bear directly on the human rat race.

Alison Mitchell

Cervical cancer

Papillomavirus and p53

Harald zur Hausen

Certain types of human papillomavirus (HPV) are linked with most cases of cervical cancer, and also with vulval, penile and perianal cancers^{1,2}. Although viral oncogenes are obviously pivotal in the development of cancer, merely expressing them is not enough — either for immortalization of cultured human cells, or for malignant conversion. Rather, additional modifications in specific cellular genes are needed, and, on page 229 of this issue, Storey and colleagues³ describe one such modification that predisposes to HPV-linked cervical cancer.

Almost ten years ago, a fresh perspective was brought on the possible mechanisms by which papillomaviruses contribute to cancer, when two HPV oncoproteins, E6 and E7, were shown to interact with two cellular proteins — p53 and retinoblastoma, respectively^{4,5}. Both p53 and retinoblastoma had previously been shown to interact with oncoproteins of other DNA tumour viruses, including SV40 and some adenoviruses. After interaction with E6, p53 was found to be degraded⁶, and, because p53 is one of the most important cellular proteins in guarding repair processes and maintaining chromosomal stability, this could explain why mutational changes are observed in human cells that are immortalized in culture by HPV.

In human populations, the p53 gene is polymorphic at amino acid 72 of the protein that it encodes — that is, p53 may contain either a proline or an arginine residue at this position. So far, no correlation has been made between either of these forms and specific human tumours (with the possible exception of lung cancer in non-smokers). But Storey *et al.* now reveal that the arginine form of p53 is more susceptible to degradation by the HPV E6 protein than is the proline form. Moreover, patients with HPV-associated cervical cancer are much more likely to contain the arginine form of p53 compared with the rest of the population. The authors conclude that patients with two copies of the arginine form have a sevenfold higher risk of developing cervical cancer than people with the proline form. Interestingly, a high percentage of skin squamous-cell carcinomas have been reported^{7,8} to contain HPV DNA, and these show an even greater prevalence of the arginine p53 modification.

If the arginine form of p53 binds more effectively to E6, it will be degraded (and hence inactivated) more rapidly, leading to increased mutation rates and chromosomal instability. Conversely, in the presence of E6, more functional activity should be maintained with the proline form of p53. Residual p53 activity has been shown in some cervical

carcinoma cell lines that contain HPV, and it would be interesting to see whether these lines contain the arginine or the proline form of p53.

The results of Storey and colleagues support the theory that cellular genes must be modified for HPV-linked carcinogenesis to occur⁹. There is likely to be a gradual accumulation of specific cellular changes during malignant progression, and evidence for this includes the long latency for tumour development after primary infection, the observed monoclonality of anogenital tumours that contain HPV, and the absence of tumour-specific modifications in the viral oncogenes. A pronounced mutator phenotype — probably mediated here by increased degradation of p53 — would facilitate accumulation of changes, increasing the risk and reducing the time required for malignant progression.

The involvement of genetic factors in HPV-linked carcinogenesis has been postulated in the past¹⁰, but careful epidemiological studies have been missing until now. In part, this is due to the high prevalence of HPV in all populations studied so far. Now that the functional consequences of the binding of E6 and E7 to polymorphic cellular proteins have been identified, we should be able to study the molecular genetics of affected populations.

In the future, other cellular genes that actively impair viral oncoproteins or suppress their transcription in non-transformed proliferating cells will probably be identified — indeed, the first candidates are already emerging². Failure or inactivation of these proteins will divert cells that carry HPV genomes down pathways to malignancy. Such discoveries will broaden our currently narrow perspective on the molecular genetics of HPV-linked cancers and, although Storey *et al.* have made an interesting and unexpected start, HPV research is still full of surprises. □

Harald zur Hausen is at the Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany.
e-mail: zurhausen@dkfz-heidelberg.de

1. IARC The Evaluation of Carcinogenic Risks in Humans Vol. 64 (Human Papillomaviruses, IARC, Lyon, 1995).
2. zur Hausen, H. *Biochim. Biophys. Acta* **1288**, F55–F78 (1996).
3. Storey, A. *et al.* *Nature* **393**, 229–234 (1998).
4. Dyson, N., Levine, A. J., Münger, K. & Harlow, E. *Science* **243**, 934–937 (1989).
5. Werness, B. A., Levine, A. J. & Howley, P. M. *Science* **248**, 76–79 (1990).
6. Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, J. M. & Howley, P. M. *Cell* **63**, 1129–1136 (1990).
7. Shumanin, V. *et al.* *J. Natl Cancer Inst.* **88**, 802–811 (1996).
8. de Villiers, E. M. *Biomed. Pharmacother.* **52**, 26–33 (1998).
9. zur Hausen, H. *Lancet* **343**, 955–957 (1994).
10. zur Hausen, H. & de Villiers, E. M. *Annu. Rev. Microbiol.* **48**, 427–447 (1994).

Daedalus

Silent flight

Combustion is a disorganized molecular process. Its product molecules fly around vigorously at random, as hot gas. A heat engine, usually a complex and inefficient piece of machinery, can extract a little of this energy as useful work.

Daedalus is now seeking a better way. In a common model of enzyme action, the reagent molecules sit neatly on the active site of the enzyme, in just the right positions to react. When the last bond of the product molecule is formed, the molecule changes shape, no longer fits, and springs off the enzyme surface, leaving the site free for further reagent molecules. This spring, says Daedalus, must represent a large part of the energy of the reaction. And it is not liberated as heat, but as molecular velocity in a specific direction. Only later, as the product molecule slams into others nearby, is it degraded to heat.

So DREADCO chemists are seeking the equivalent phenomenon on surfaces that catalyse combustion. A well-ordered surface might well fire off the final CO₂ or H₂O molecules vigorously and directedly, as they take their final shape and cease to fit the surface. In a low-density gas environment, they would travel some distance before hitting any other molecule. The catalytic surface would feel the full reaction force of their departure.

The obvious application is to aircraft propulsion. So the researchers are taking their most promising catalytic materials, and vacuum-evaporating them onto model gliders. The evaporation directions are cunningly chosen so that the deposited catalytic crystallites acquire exactly the right orientation to the local wing surface. Each glider is then being flown in a dilute, non-explosive mixture of air with a fuel gas, such as hydrogen or ethylene. With the right catalyst, the products of surface combustion will not be disengaged as hot gas, but will be ejected directionally downwards and backwards. Their reaction will give lift and thrust. In its novel atmosphere, the glider will speed and soar indefinitely.

A realistic aircraft, however, has to fly in pure air on liquid fuel. The fuel will be pumped through porous regions on the catalytic surfaces, to diffuse over them as a monolayer. The new craft will be propelled by its whole surface. It will be utterly silent, free of heavy and expensive engines, and powered with greater than Carnot efficiency. It will transform aviation. Sadly, its owners will have to forego the usual tasteless painted colour-schemes.

David Jones

Obituary

Jean Rouxel (1935–98)

Pioneer in *chimie douce*

Jean Rouxel, who died suddenly on 19 March aged 63, had only a few months previously received the highest award that France can bestow on a scientist — the annual gold medal of the CNRS, the Centre National de la Recherche Scientifique. Rouxel was primarily a solid-state chemist, but had a wide-ranging influence on all aspects of inorganic chemistry and materials science. His main standing was as one of the founding fathers of *chimie douce* ('soft chemistry').

Until the 1960s, solid-state chemistry was a 'hard chemistry' that involved the use of very extreme conditions — principally high temperatures — for the synthesis of compounds; such conditions are necessary to overcome the kinetic barriers to diffusion and reactivity, and lead to a unique materials phase that is stable in the chosen conditions. *Chimie douce*, by contrast, is carried out at low temperature and allows the synthesis of very many atomic arrangements; these are often metastable but have a large variety of structures and properties.

Rouxel initiated a part of *chimie douce* devoted to intercalation–deintercalation processes with low-dimensional solids. Such solids can be regarded as stacks of infinite one-dimensional or two-dimensional molecules, in chains or sheets respectively, and they constitute a fruitful intermediate between molecular inorganic chemistry and solid-state chemistry.

The approach resulted in the discovery of many new compounds, especially among the chalcogenides which are composed of a chalcogen, such as sulphur, selenium or tellurium, and a more metallic element. But it also revealed new concepts, physical effects and chemical properties, and has applications in fields as various as battery design, catalysis and electrochromism (in which pigments change colour when an electric field is applied, the compounds concerned being used in displays).

Rouxel was born in 1935 in a small village in Brittany, and throughout his life remained strongly influenced by Breton customs and culture. He obtained his early scientific education in Brest and later at the University of Rennes, where he met a young, newly appointed professor in inorganic chemistry, Paul Hagenmuller. It was with Hagenmuller that Rouxel discovered and embraced with eagerness the joys — and the attendant doubts — of a research career. Appointed an assistant lecturer in 1957, a few years later he

followed Hagenmuller and his group to Bordeaux, obtaining his PhD there in 1961.

At that time, Nantes became an independent university and appointed Rouxel a professor in 1963. There he created his own laboratory, gathering a cadre of young and energetic colleagues around him. Fifteen years later, the group received the support of the CNRS, attaining the status of associate laboratory; finally, in 1988, the Institute of Materials at Nantes was created with Rouxel as head. This institute was and remains a model of interdisciplinarity, where solid-state and molecular chemists, solid-state physicists and other specialists in materials science all rub shoulders.

Rouxel's scientific achievements stemmed from the originality of his work on chemical bonding in low-dimensional solids, materials which he manipulated as if they were sheets of paper in a book between which he could insert various molecules. Most usually, the molecules concerned are cations, which donate electrons to the sheet and thereby reduce it. This reduction of the sheet by increasing the number of *d* electrons per metal atom sometimes leads to metal–metal bonding (like that observed in clusters), and sometimes to electronic delocalization in bands (like that seen in metals). Occasionally an intermediate situation occurs, as it does in niobium selenide, with the formation of charge density waves. Rouxel viewed these in a fresh light, as bonding density waves, thereby showing that physicists and chemists are often looking at the same phenomena with complementary eyes.

The reverse process, deintercalation of cations out of the sheets, of course involves oxidation and is one of the most powerful synthesis processes in *chimie douce*. But it raises some difficult questions. For instance, from which band do the electrons that are lost from the sheet come — the metallic *d* orbitals or the *sp* anionic bands? Rouxel showed that what matters is the relative energies of the electron bands. For example, with some pairings of metal and chalcogen, such as iridium with tellurium, it is

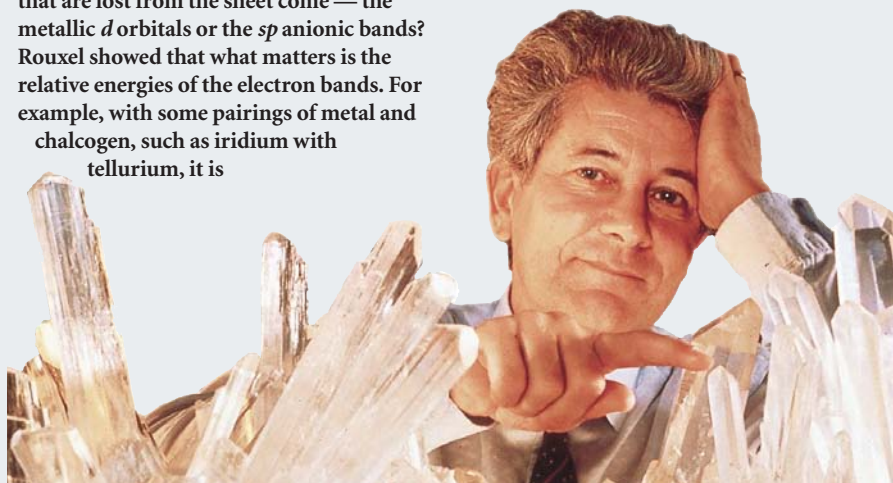
possible to empty the *p* anionic band first; this creates holes in place of the antibonding electrons, which stabilizes the tellurium–tellurium anionic bonds. This is equivalent to adding bonding electrons to an empty band — both promote bonding. This chemistry of antibonding electrons was the final area to be developed by Rouxel during his highly fruitful career.

Before that, he had tackled another fundamental question to do with so-called 'misfit' structures. These are two-dimensional solids with two kinds of sheet but only one common crystallographic parameter, a peculiar property which marks a conceptual limit to the basic notion of a thermodynamically well-defined phase.

Jean had an endearing personality, being both modest and down-to-earth, yet capable of great enthusiasm — he took the same intense interest in the roses in his garden as in being an ambassador for French science at international meetings. He also always liked to have young scientists around him, and closely followed their scientific upbringing. During the 1970s, he created a summer school where the oldest students taught the youngest ones the basic concepts and methods of solid-state chemistry, informally and in the friendliest of atmospheres. The first meeting took place in Brittany, not too far from Nantes, and was named after a nearby, moorland village, Galerne. A saying of the local farmers, "The wind is from Galerne", means that an invigorating wind is blowing from the west, from the direction of the steeple of Galerne's church. One might say the same of Jean Rouxel — his was an invigorating influence on solid-state chemistry that blew steadily from the west of France.

Michel Pouchard

Michel Pouchard is in the Institut de Chimie de la Matière Condensée de Bordeaux – CNRS, Avenue du Dr A. Schweitzer, 33608 Pessac Cedex, France. e-mail: pouchard@icmcb.u-bordeaux.fr



DE LAUBIER-FIGARO/FRANK SPOONER

Hyde's horrors

Can you tell a criminal from the look of his face? The "downright detestable" appearance of Robert Louis Stevenson's evil Mr Hyde stands in the same tradition as the images used by Darwin in his work on pathognomics.

NATIONAL FILM ARCHIVE/ROBAL COLLECTION



Spencer Tracy as Dr Jekyll and Mr Hyde from Victor Fleming's 1941 film.

Martin Kemp

How the image of science and its practitioners enters public consciousness — frequently in caricatured form — cannot be dismissed as irrelevant to its professional practice. Science institutions are as subject to political decision-taking in the context of public opinion as any other bodies funded from state and private purses. And even the grossest of visual distortions feed on images promulgated by scientists themselves. The 'horror' film is a case in point.

Robert Louis Stevenson, the Edinburgh engineer and qualified lawyer, wrote his classic *The Strange Case of Dr Jekyll and Mr Hyde* (1886) in the heyday of physiognomics, craniology and anthropometry, and in the wake of Francis Galton's new science of eugenics. These disciplines flourished in a scientific climate that was obsessed with the classification and definition of 'types' through precise characterization and minute measurement. In Stevenson's original text, it is striking how subtle and elusive were the visual signs of the degenerate and regressive Hyde released by Jekyll's chemical cocktail.

Mr Enfield, the first witness, typifies the problems: "He is not easy to describe. There is something wrong with his appearance; something displeasing, something downright detestable. I never saw a man so disliked, and yet I scarce know why. He must be deformed somewhere; he gives a strong feeling of deformity, although I couldn't specify the point."

A conviction that measurements of extraordinary refinement — beyond simple descriptions — were needed to define 'types' through facial and cranial formations lay at the heart of several large-scale research enterprises in the nineteenth century.

These included the programme of craniometry undertaken by Paul Broca, founder of the Parisian Anthropological Society in 1859, the photographic databanks of eugenic groups built up by Galton in the 1870s and 1880s, and Alphonse Bertillon's programme of criminological measurement undertaken for the Paris police towards the end of the century.

Inevitably, when "the animal within me licking the chops of memory" (in Jekyll's vivid evocation) came to be visualized in popular cinema, a more blatant characterization was required than that in the novel. The physiognomy provided by director Victor Fleming for Spencer Tracy in 1941 was

not the most crudely bestial of the filmed Hydes, but his range of manic expressions stood in a long tradition of exaggerated pathognomics. The images used by Charles Darwin in his *The Expression of the Emotions in Man and Animals* (1872), including the remarkable photographs of Duchenne de Boulogne and drawings by leading illustrators, stand in this tradition.

Darwin set human expression in the broader context of evolution, not in terms of traditional parallels between human facial types and animal physiognomies but through an understanding of the origins of the expressive mechanisms.

He wrote: "As long as man and all other animals are viewed as independent creations, an effectual stop is put to our natural desire to investigate as far as possible the causes of expression.... With mankind some expressions, such as the bristling of the hair under the influence of extreme terror, or the uncovering of the teeth under that of furious rage, can hardly be understood except on the belief that man once existed in a much lower and animal-like condition."

We still talk of those who perpetrate atrocities as 'animals' or 'beasts', and scrutinize the faces of criminals that stare out of newspaper photographs for signs of bestiality. And the Galton laboratory at University College London has a web site seeking physiognomic volunteers for its collaborative project on the genetics of human facial features, for which the Forensic Science Service is a funding agency. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

e-mail: martin.kemp@trinity.oxford.ac.uk



Fig. 17. The same, when pleased by being caressed.

Joseph Wolf's Nger macaque (*Cynopithecus niger*) aroused by stroking, engraved in Charles Darwin's *The Expression of the Emotions in Man and Animals* (1872).

WELLCOME INSTITUTE LIBRARY, LONDON

Human gene for physical performance

A specific genetic factor that strongly influences human physical performance has not so far been reported, but here we show that a polymorphism in the gene encoding angiotensin-converting enzyme does just that. An 'insertion' allele of the gene is associated with elite endurance performance among high-altitude mountaineers. Also, after physical training, repetitive weight-lifting is improved eleven-fold in individuals homozygous for the 'insertion' allele compared with those homozygous for the 'deletion' allele.

The endocrine renin-angiotensin system is important in controlling the circulatory system. Angiotensin-converting enzyme (ACE, or kininase II) degrades vasodilator kinins, and converts angiotensin I (ATI) to the vasoconstrictor angiotensin II (ATII). In addition, local renin-angiotensin systems may influence tissue growth¹. A polymorphism of the human ACE gene has been described in which the deletion (*D*) rather than insertion (*I*) allele is associated with higher activity by tissue ACE².

There is evidence for a skeletal muscle renin-angiotensin system³, suggesting that muscle growth, and thus physical performance, might be positively associated with the *D* allele. However, our initial studies suggested that the *I* allele was associated with improved endurance performance. We investigated this association in two parallel experiments.

High-altitude mountaineers perform extreme-endurance exercise. Thirty-three elite unrelated male British mountaineers, with a history of ascending beyond 7,000 metres without using supplementary oxygen, were identified by the British Mountaineering Council. DNA was extracted from a mouthwash sample of the 25 male respondents, and ACE genotype was determined using a three-primer polymerase chain reaction amplification⁴.

Genotype distribution was compared with that of 1,906 British males free from clinical cardiovascular disease⁵. Mean age was 40.6 ± 6.5 years in the 25 subjects, and 55.6 ± 3.2 years for the controls. Both groups were in Hardy-Weinberg equilibrium. Both genotype distribution and allele frequency differed significantly between climbers and controls (Fig. 1a; P was 0.02 and 0.003 respectively (χ^2 test)), with a relative excess of *II* genotype and deficiency of *DD* genotype found in the climbers.

Among the 15 climbers who had ascended beyond 8,000 m without oxygen, none was homozygous for *D* (6 *II* and 9 *ID*: *I* allele frequency = 0.65). Further, ranked by number of ascents above 8,000 m without oxygen, the top performer was homozygous

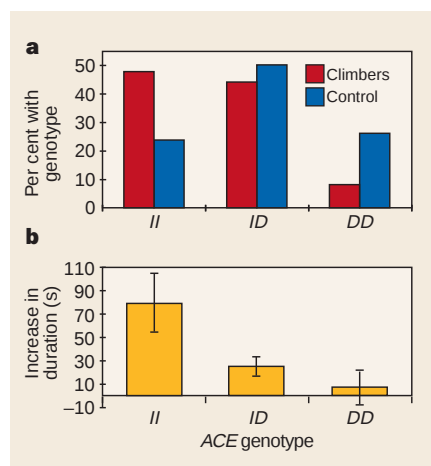


Figure 1 Association of ACE alleles with performance. **a**, Distribution of ACE *II/D* genotypes in 25 elite British mountaineers and 1,906 healthy British men. **b**, Mean ($\pm$ s.e.m.) of individual improvement in duration of repetitive elbow flexion after 10 weeks of physical training among British army recruits. An initial cycle frequency of 0.3 Hz was strictly regulated by an electronic metronome and was increased by 0.05 Hz every minute. Performance was independent of skill or variation in technique. For each individual, the absolute difference between pre- and post-training data was obtained, and the mean of these differences calculated. Pre-training and post-training data were compared using two-tailed paired *t*-tests. Means were compared by generalized linear modelling with Statistical Analysis Software¹³. Data were adjusted by including baseline values as a covariate in the statistical model.

for *I* (5 ascents, compared with a mean of 2.4 ± 0.3 ascents, or 1.44 ± 0.3), as were the top two in this group for number of additional 7,000-m ascents (> 100 and 18, compared with a mean of 10.3 ± 6.5 ascents).

In a second study, ACE genotype was determined in 123 Caucasian males recruited to the UK army consecutively. Seventy-eight completed an identical 10-week general physical training programme (age, 19.0 ± 0.2 years; height, 176.6 ± 0.7 cm; body mass index, 22.2 ± 0.2 kg m⁻²). Their ACE genotype (20 (25.6%) *II*, 46 (59.0%) *ID*, 12 (15.4%) *DD*) matched that of those who failed training, as did their physical characteristics (neck, chest and waist circumference, elbow diameter and armspan), and all characteristics were independent of genotype.

The maximum duration (in seconds) for which they could perform repetitive elbow flexion while holding a 15-kg barbell was assessed both before and after the training period. Pre-training performance was independent of genotype (mean, 119.8 ± 6.2 s). Duration of exercise improved significantly for those (66 individuals) of *II* and *ID* geno-

type (79.4 ± 25.2 and 24.7 ± 8.8 s: P was 0.005 and 0.007 respectively) but not for the 12 of *DD* genotype (7.1 ± 14.9 s: $P = 0.642$) (Fig 1b). Improvement was thus eleven-fold greater ($P = 0.001$) for those of *II* than for those of *DD* genotype.

Genotype-dependent improvements were unlikely to be due to changes in individual muscle fibre size and strength (which need more than three months of specific strength-training to occur) or altered co-ordination, neural firing pattern or recruitment of fast motor units (given the lack of specific training for the test task)⁶⁻⁸. Increased performance is therefore most likely to be due to an improvement in the endurance characteristics of the tested muscles.

The association of the *I* allele with improved endurance might derive from variable increases in substrate delivery due to increases in cardiac output and muscle capillary density; from changes in the nature of substrate used, due to a differential shift to stored fatty acids as fuel⁹, or in the efficiency of substrate utilization relating to altered muscle fibre type; from altered mitochondrial density, or from raised muscle myoglobin content^{10,11}. Elevated catecholamine, cortisol and growth hormone concentrations may all increase the availability of oxidative fuel¹².

Further work will be needed to determine whether this correlation holds beyond the limited group studied here and to explore the mechanisms underlying these observations.

**H. E. Montgomery, R. Marshall,
H. Hemingway, S. Myerson**

University College London Centre for Cardiovascular Research, Rayne Institute, University Street, London WC1E 6JJ, UK

P. Clarkson, C. Dollery, M. Hayward

University College London Hospital, London WC1E 6DB, UK

D. E. Holliman

Centre for Human Sciences, DERA Farnborough GU14 0LX, UK

M. Jubbs, M. World

Royal Defence Medical College, HMS Dolphin, Gosport, Hampshire PO12 2AB, UK

E. L. Thomas, A. E. Brynes, N. Saeed,

M. Barnard, J. D. Bell

Imperial College School of Medicine MRI Unit, London W12 0HS, UK

K. Prasad

St George's Hospital, Department of Cardiovascular Sciences, London SW17 0RE, UK

M. Rayson

Optimal Performance, Farnham GU9 7EB, UK

P. J. Talmud, S. E. Humphries

University College London Centre for Cardiovascular Genetics, Rayne Institute, University Street, London WC1E 6JJ, UK

1. Katz, A. M. *Heart Disease and Stroke* 1, 151–154 (1992).
2. Danser, A. H. *et al. Circulation* 92, 1387–1388 (1995).
3. Reneland, R. & Lithell, H. *Scand. J. Clin. Lab. Invest.* 54, 105–111 (1994).
4. Montgomery, H. E. *et al. Circulation* 96, 741–747 (1997).
5. Miller, G. J., Bauer, K. A., Barzegar, S., Cooper, J. A. & Rosenberg, R. D. *Thromb. Haemost.* 75, 767–771 (1996).
6. Komi, P. V. in *Biochemistry of Exercise VI* (ed. Saltin, B.) 529–575 (Human Kinetics, Champaign, Illinois, 1986).
7. Rutherford, O. M., Greig, C. A., Sargeant, A. J. & Jones, D. A. *J. Sports Sci.* 4, 101–107 (1986).
8. Jones, D. A. & Rutherford, O. M. *J. Physiol.* 391, 1–11 (1987).
9. Rennie, M. J., Winder, W. W. & Holloszy, O. *Biochemistry* 156, 647–655 (1976).
10. Bloom, S. R., Johnson, R. H., Park, D. M., Rennie, M. J. & Salaiman, W. R. *J. Physiol.* 258, 1–18 (1976).
11. Hudlicka, O. *News Physiol. Sci.* 3, 117–120 (1988).
12. Wasserman, D. H. & Vranic, M. in *Biochemistry of Exercise VI* (ed. Saltin, B.) 167–179 (Human Kinetics, Champaign, Illinois, 1986).
13. SAS Institute. *SAS Users Guide* (SAS Institute, Cary, North Carolina, 1985).

Meteoritic oxide grain from supernova found

Meteorites contain tiny (0.002–10 µm) mineral grains which formed around stars or in stellar explosions^{1,2}. These grains have unusual isotope compositions that reflect those of the stars in which they formed. Stardust composed of nano-diamonds, silicon carbide (SiC), silicon nitride (Si₃N₄) or graphite are believed to derive from a range of stellar types^{1,2}, whereas the oxygen-rich grains found to date are thought to originate only in red giants and asymptotic giant branch stars³. We report here an oxide grain that is extremely rich in the isotope oxygen-16 (¹⁶O), in an acid-resistant residue of the Tieschitz meteorite. This grain, T84, probably derives from the ejecta of a type II supernova and is the first reported oxide grain derived from such a source, despite ¹⁶O being the third most abundant isotope

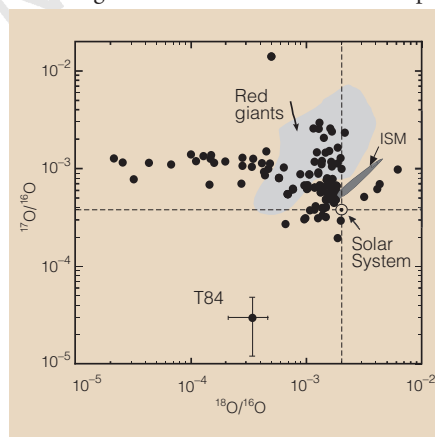


Figure 1 Oxygen-isotopic ratios of meteoritic grains. Filled circles, 103 presolar oxide grains from meteorites; light grey, range observed in red giants; dark grey, range in the gaseous interstellar medium (ISM). Grain T84 is unusually rich in ¹⁶O, and probably formed in the ejecta of a type II supernova.

ejected by supernovae (after hydrogen and helium)⁴.

The favoured explanation for the origin of ¹⁶O enrichments in calcium–aluminium-rich inclusions (CAIs) in meteorites has been that ¹⁶O was carried into the solar nebula by refractory oxide grains (for example, aluminium oxide, Al₂O₃) from supernovae⁵. Although our discovery of T84 indicates that micrometre-sized refractory oxide grains do indeed form in supernova ejecta, the abundance of such grains in meteorites is probably too low to account for the ¹⁶O excesses observed in CAIs.

We compare the oxygen-isotope ratios of T84 with those of 102 other presolar oxide grains from meteorites³, together with the ranges of these ratios in red giants⁶ and in the interstellar medium⁷ (Fig. 1). Like most of the grains, T84 was identified by isotopic imaging in an ion microprobe³. Its isotopic composition is: ¹⁷O/¹⁶O = $3.0 \pm 1.8 \times 10^{-5}$, ¹⁸O/¹⁶O = $3.4 \pm 1.3 \times 10^{-4}$.

Unfortunately, T84 was completely consumed during measurement because of its small size (about 0.5 µm) and its mineralogy cannot now be determined. We estimate that the abundance of grains such as T84 in Tieschitz is less than 0.25 parts per billion.

The composition of T84 indicates that it formed in a stellar environment very rich in ¹⁶O. Massive stars synthesize copious amounts of ¹⁶O by burning helium, carbon and neon deep in their interiors⁴. The freshly made ¹⁶O is expelled into the interstellar medium either by type II supernova explosions or, in stars whose masses are more than 30 times that of the Sun, by strong winds during a so-called WO phase⁸. Type Ia supernovae also eject material rich in ¹⁶O (ref. 9).

Either supernovae or WO stars are potential sources of T84, but WO stars are extremely rare in the Galaxy whereas supernova explosions are relatively common. Furthermore, type II supernovae eject far more ¹⁶O than do type Ia, so a type II origin for T84 is more likely.

The discovery of T84 shows that oxide grains from supernovae were present in the early Solar System, but the abundance of such grains in meteorites is surprisingly small. Galactic dust production rates, calculated supernova yields, and the observed fraction (1%) of presolar SiC grains from supernovae suggest that Al₂O₃ grains produced by supernovae should make up 10–70% of presolar oxides^{3,10} rather than the amount observed, which is less than 1%.

The dust produced by supernova 1987A (ref. 11) is much smaller than the grains described here and T84 is at the small end of what can be currently analysed in the ion microprobe. Supernovae may thus produce much smaller

oxide grains than do red giants, and they could have been systematically excluded from the presolar oxide data set. Furthermore, detailed analysis of SN1987A's spectra indicates that the dust formed there was not oxygen-rich, but was possibly metallic iron¹¹. Perhaps oxides do not easily condense from supernova ejecta, despite the extremely oxidizing conditions.

The very low abundance of presolar ¹⁶O-rich oxide grains seems to rule out supernova Al₂O₃ as the prime carrier of ¹⁶O enrichments in CAIs. Instead, ¹⁶O may have been carried into the early Solar System by supernovae-derived ¹⁶O-rich silicates rather than oxides. Alternatively, the Solar System's oxygen-isotope heterogeneity may have resulted largely from non-mass-dependent isotopic fractionation¹².

L. R. Nittler, C. M. O'D. Alexander, J. Wang

Department of Terrestrial Magnetism,
Carnegie Institution of Washington,
5241 Broad Branch Road NW, Washington,
DC 20015, USA

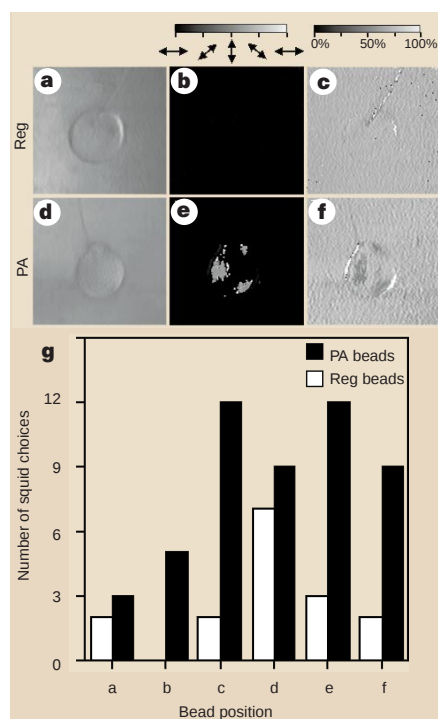
X. Gao

McDonnell Center for the Space Sciences,
Washington University, 1 Brookings Drive,
St Louis, Missouri 63130, USA

1. Anders, E. & Zinner, E. *Meteoritics* 28, 490–514 (1993).
2. Zinner, E. in *Astrophysical Implications of the Laboratory Study of Presolar Materials* (eds Bernatowicz, T. J. & Zinner, E.), No. 402 in AIP Conf. Proc. 3–26 (AIP, Woodbury, New York, 1997).
3. Nittler, L. R., Alexander, C. M. O'D., Gao, X., Walker, R. M. & Zinner, E. *Astrophys. J.* 483, 475–495 (1997).
4. Woosley, S. E. & Weaver, T. A. *Astrophys. J. Suppl.* 101, 181–235 (1995).
5. Clayton, R. N., Grossman, L. & Mayeda, T. K. *Science* 182, 485–488 (1973).
6. Smith, V. V. & Lambert, D. L. *Astrophys. J. Suppl.* 72, 387–416 (1990).
7. Wilson, T. L. & Rood, R. T. *Annu. Rev. Astron. Astrophys.* 32, 191–226 (1994).
8. Prantzos, N., Doom, C., Arnould, M. & de Loore, C. *Astrophys. J.* 304, 695–712 (1986).
9. Thielemann, F. K., Nomoto, K. & Yokoi, K. *Astron. Astrophys.* 158, 17–33 (1986).
10. Alexander, C. M. O'D. in *Astrophysical Implications of the Laboratory Study of Presolar Materials* (eds Bernatowicz, T. J. & Zinner, E.), No. 402 in AIP Conf. Proc. 567–594 (AIP, Woodbury, New York, 1997).
11. Wooden, D. in *Astrophysical Implications of the Laboratory Study of Presolar Materials* (eds Bernatowicz, T. J. & Zinner, E.), No. 402 in AIP Conf. Proc. 317–378 (AIP, Woodbury, New York, 1997).
12. Thieme, M. H. & Heidenreich, J. E. *Science* 219, 1073–1075 (1983).

Polarization vision helps detect transparent prey

Transparency enables aquatic organisms to avoid detection by visual predators. But we have found that this camouflage can be broken using a visual mode evolved by several predators, such as squid. Under partially linearly polarized lighting, squid detect zooplankton prey at a distance 70% greater than those achieved under non-polarized illumination. The role of polarization sensi-



tivity in predation is confirmed by squid's preference for transparent, yet polarization-active, targets that mimic their prey.

In aquatic environments, many organisms are transparent or translucent, rendering them difficult to detect¹. However, some of these planktonic organisms contain polarization-active tissues².

Using the new retardance-measuring polarized-light microscope (Pol-Scope)³ and standard linear polarization microscopy, we inspected the tissues of 20 zooplanktonic species to determine the polarization-active structures in their bodies. Muscle fibres and rhabdomeric photoreceptors showed strong retardance up to one-quarter of the wavelength. Edge birefringence could be detected in only a few organisms, and was most noticeable in chaetognaths and in the antennae of crustaceans.

Natural illumination is often partially linearly polarized⁴, so transparent but polarization-active organisms may be rendered conspicuous when viewed by a polarization-sensitive visual system⁵. We tested the role of polarization vision in prey selection and detection in adults and hatchlings of the squid *Loligo pealei*, which is known to be polarization-sensitive⁶.

We examined the choices of adult squid when presented with small transparent glass beads that mimicked prey items and which were either polarization-active (PA) or had no effect on the polarization characteristics of light passing through them (regular beads, Reg). Six of these beads (three Reg, three PA) were suspended in mid-water at one end of a rectangular tank measuring 370 × 120 cm and containing water to a depth of 40 cm. Of 50 animals tested,

Figure 1 Adult squid choices of glass beads that mimicked prey. **a–f**, Glass beads (1 cm diameter, equal transparency in the 400–650 nm range) appear identical to a human observer (**a,d**). Using a submersible imaging polarimeter⁹, the light intensity (**a,d**), e-vector orientation (**b,e**) and percentage of linear polarization (**c,f**) were measured at each pixel in an image. Some beads (**d–f**) were heat-stressed during manufacture and hence were polarization-active (PA), whereas others (**a–c**) were not stressed and did not affect the light's polarization (Reg). **g**, Fifteen squid examined more than one bead in the 15-minute period, resulting in 66 bead choices. Scoring was by one of two volunteers, neither of whom knew the identities (Reg or PA) of the glass beads. Bead position (**a–f**, across the tank, 20 cm from each side, and 16 cm between them; **a**, left-most position as a squid approached the beads) was a significant factor in the animals' choices ($\chi^2 = 11.1$, $P < 0.05$), but there was a significant preference for the PA beads independent of the bead's location ($\chi^2 = 11.22$, $P < 0.05$).

10 did not respond to the beads and were excluded from analysis.

Of the remaining 40, half were tested with the PA beads in positions as shown in Fig. 1a, c, e and half with the PA beads as in Fig. 1b, d, f. Beads were illuminated from behind through a glass window. The light was filtered through a wax-paper depolarizer and a Polaroid HN38S linear polarizing filter, creating a partially polarized light field comparable to the marine environment⁴. Additional dim illumination was provided by overhead fluorescent lights. Results (number of bead choices) were analysed using the χ^2 goodness-of-fit test with equal expected frequencies.

Adult squid preferred PA beads over Reg beads (Fig. 1g). This preference was evident from: total squid choice in a 15-min period (Fig. 1; PA = 50, Reg = 16; $\chi^2 = 17.52$, $P < 0.0001$); first approach of each squid (PA = 33, Reg = 7; $\chi^2 = 16.90$, $P < 0.0001$); and the squids' overall preferences (PA = 30, Reg = 4, equal number of Reg and PA choices = 6; $\chi^2 = 19.88$, $P < 0.0001$).

We also examined predation on live zooplankton by squid hatchlings. We introduced hatchlings ($n = 250$) into an experimental tank illuminated through a side window with either linearly polarized or depolarized light. We recorded the hatchlings' feeding behaviour with a video camera for 60 minutes. The first 100 attacks in each trial (polarized or depolarized illumination) were measured when the predator:prey density was 1:50. We also measured 25 attacks in each illumination setting with a predator:prey ratio of 1:140.

Squid hatchlings attacked planktonic prey under polarized illumination at a 70% greater distance than under depolarized illu-

mination (6.09 ± 2.92 , as against 3.58 ± 1.26 squid body lengths, average $\pm$ s.d., $P < 0.001$, Kolmogorov–Smirnov 2-sample test). Prey densities did not affect the attack distances ($P > 0.1$, Kolmogorov–Smirnov 2-sample test). Attacks had similar success rates under polarized and depolarized illumination (86 and 74%, $\chi^2 = 7.78$, $P > 0.05$, χ^2 with equal expected frequencies), indicating that a greater attack distance does not impair prey capture.

Loligo pealei, like other squid, is a pelagic visual predator. Both young and adults feed on transparent zooplankton^{7,8}. Adults are predominantly crepuscular predators, feeding at times when the partial polarization of the underwater light field is at its maximum⁷. We predict that polarization-based predation may be common among other polarization-sensitive planktivores such as fish and crustaceans.

Nadav Shashar, Roger T. Hanlon

Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA

Anne deM. Petz

Wellesley College, 106 Central Street, Wellesley, Massachusetts 02181, USA

e-mail: nshashar@mbl.edu

1. Nilsson, D. E. In *Zooplankton: Sensory Ecology and Physiology* (eds Lenz, P. H., Hartline, D. K., Purcell, J. E. & Macmillan, D. L.) 149–162 (Gordon and Breach, Amsterdam, 1996).
2. Hull, P. G., Hunt, A. J., Quinby-Hunt, M. S. & Shapiro, D. B. *Underwater Imaging, Photography, and Visibility. Soc. Photo-Optical Instrumentation Engineers Proc. Series 1537*, 21–29 (1991).
3. Oldenbourg, R. & Mei, G. J. *Microsc.* **180**, 140–147 (1995).
4. Waterman, T. H. & Westell, W. E. *J. Mar. Res.* **15**, 149–169 (1956).
5. Fineran, B. A. & Nicol, J. A. C. *Phil. Trans. R. Soc. Lond. B* **283**, 25–60 (1978).
6. Saidel, W. M., Lettvin, J. Y. & McNichol, E. F. *Nature* **304**, 534–536 (1983).
7. Vovk, A. N. *Les. Trudy Atl. Nauchno-Issled. Inst. Rybn. Khoz. Okeanogr.* **42**, 141–151 (1974).
8. Hanlon, R. T., Turk, P. E., Lee, P. G. & Yang, W. T. *Fish. Bull.* **85**, 163–167 (1987).
9. Cronin, T. W., Shashar, N. & Wolff, L. Portable imaging polarimeters. *Proc. 12th Int. Congr. Pattern Recognition* 606–609 (1994).

Ultradry seed storage cuts cost of gene bank

Today we are faced with an urgent challenge to organize the efficient conservation of biodiversity. In the case of plants, one way of doing this is to store their seeds in a gene bank. The effective preservation of seeds depends on their moisture content and on the storage temperature, which can present a problem in developing countries where the costs of cold storage are prohibitive. Here we describe a way of storing seeds under ultradry conditions at ambient temperatures which not only improves their longevity and vigour but also promises to be a cost-effective technique for germplasm conservation.



Figure 1 The effects of ultradry storage on *Eucommia* seeds. No difference in viability or vigour was evident in seeds stored for two years either at ambient temperature after ultradrying or at low temperature. **a**, Ultradried seeds stored at ambient temperature: moisture content (m.c.) was 3.3%, germination was 97%, vigour index was 26.7. **b**, Seeds stored under conventional conditions at -20°C : m.c. was 8.4%, germination was 96%, vigour index was 24.1. **c**, Control stored at ambient temperature: m.c. was 7.5%, viability was lost.

Seed moisture content and storage temperature are the most important factors affecting seed longevity during storage^{1,2}, with the most commonly used conditions in gene banks being 5–7% or 5–6% (refs 2 and 3) moisture content at -10°C to -20°C . Previous work⁴ has shown that low moisture content can substitute for low temperature in helping to preserve some orthodox seeds. We reasoned that if this substitution were to apply to a wide range of plant species, then refrigeration could be reduced or omitted by controlling the seeds' moisture content and storing them at ambient temperature, making the idea of a low-cost gene bank of ultradried seeds feasible.

We conducted a series of experiments on ultradry seeds (less than 5% moisture content) at the Beijing Botanical Gardens over a period of nine years. More than 20 species of seeds were tested, including oil seeds, starchy seeds and some seeds rich in protein: examples were *Arachis hypogaea*, *Asparagus officinalis*, *Beta vulgaris*, *Brassica campestris*, *B. oleracea*, *B. pekinensis*, *Carthamus tinctorius*, *Cucumis sativus*, *Cucurbita pepo*, *Eucommia ulmoides*, *Glycine max*, *Oenothera lamarckiana*, *Oryza sativa*, *Populus nigra*, *Raphanus sativus*, *Sesamum indicum*, *Sorghum vulgare*, *Triticum aestivum*, *Ulmus macrocarpa* and *U. pumila*.

First we tested whether the ultradry conditions themselves affected the seeds and found that in general they were harmless, apart from a few exceptions such as some cultivated varieties of Chinese

soybean produced by selective breeding. We found that our ultradry conditions did not induce any significant change either in the proportion of seeds that germinated or in their vigour⁵.

Moreover, these ultradried seeds aged more slowly and tolerated storage at ambient temperatures better than controls^{5,6}. The overall effect on the seeds was much the same as that of cold (-20°C) (Fig. 1) or ultracold (-196°C) storage. Experimental data confirmed that the ultradry seeds maintain their genetic stability after storage or accelerated ageing^{7,8}.

Improvements compared with controls were seen for various parameters, including the botanical features of the seedlings, the proportion of chromosome aberrations, integrity of isoenzymes and survival of DNA polymorphisms. According to different physiological, biochemical and cytological criteria, storage of seeds under our ultradry conditions had a positive effect on seedling performance, photosynthesis efficiency and respiration intensity; on membrane function, self-defence against lipid peroxidation and the products of deterioration; and on the ultrastructure of radicle cells, as seen in the transmission electron microscope, and of cotyledons, as seen in the scanning electron microscope.

Three factors need to be taken into account before the ultradry seed storage technique can be used. First, the optimal moisture content must be determined; second, damage to particularly sensitive types of orthodox seeds must be prevented; and third, an effective pretreatment needs to be

designed to ensure satisfactory germination of ultradried seeds.

Tests have indicated that the most suitable upper and lower limits of moisture content depend on the type of seed⁹. Thermodynamic analysis of the seed water may help in predicting a seed's tolerance to desiccation for some ultradried seed types, but accurate measurement of moisture content may also be necessary for others¹⁰.

The seeds of soybean (some Chinese cultivars) and rice (hybrid) were sensitive to the ultradrying process. Pretreatment by hydration–dehydration before drying is one of the most effective ways to protect them against damage, and works by altering the thermodynamic parameters. Some antioxidants also improved the storability of ultradried seeds.

Imbibitional injury often occurs to seeds during germination after ultradrying and storage, especially to legume seeds. Consequently, before germination or field planting of such seeds, their moisture content should be increased to more than 8% by using an equilibrium premoistening method based on osmoconditioning principles¹¹.

Ultradry seed storage is simple, easy and inexpensive. Seeds can be dried in a desiccator until the correct moisture content is reached (silica gel or quicklime are ideal desiccants); they can then be stored in a sealed container at ambient temperature. Our results demonstrate that preparing and storing seeds in this way could be a useful and cost-effective technique for germplasm conservation, particularly in developing countries, because of its low capital and operational costs.

Guang-Hua Zheng, Xin-Ming Jing

Beijing Botanical Garden, Institute of Botany,
CAS, Beijing 100093, China
e-mail: ydzheng@netra.nju.edu.cn

Kar-Ling Tao

Food and Agriculture Organization of the United Nations, AGPS (Seed and Plant Genetic Resources Service), 00100 Rome, Italy

- Justice, O. L. & Bass, L. N. in *Principles and Practice of Seed Storage* 26 (Agriculture Handbook No. 506, Castle-House, Beccles & London, 1979).
- Zheng, G. H., Shi, Z. L., Zhao, T. F. & Tao, K. L. in *The Practical Seed Physiology* 371–423 (Agriculture Publishing House, Beijing, 1990).
- Paroda, R. C. & Arora, R. K. in *Plant Genetic Resources Conservation and Management* 16–18 (IBPGR Regional Office, Malhotra Publishing House, New Delhi, 1991).
- Zheng, G. H. *Scientia Agricultura Sinica* **13**, 37–43 (1980).
- Zheng, G. H. *Chinese Biodiversity* **2** (suppl.), 61–65 (1994).
- Cheng, H. Y., Zheng, G. H. & Tao, K. L. *Plant Genetic Resources Newslett.* **83/84**, 9–14 (1991).
- Jing X. M., Zheng G. H. & Tao K. L. *Acta Botanica Sinica* **36** (suppl.), 73–78 (1994).
- Cheng, H. Y. et al. *Plant Genetic Resources Newslett.* **110**, 35–40 (1997).
- Lin, J., Zheng, G. H. & Jing, X. M. *Plant Introduction and Acclimatization* **11**, 104–110 (1997).
- Cheng, H. Y., Zheng, G. H., Qin, H., Tao, K. L. & Zhou, M. D. *Scientia Agricultura Sinica* **29**, 65–73 (1996).
- Zheng, G. H. *Seed Sci. Res.* **1**, 27–134 (1991).

China's long march into the space age

The Chinese Space Programme: From Conception to Future Capabilities

by Brian Harvey

Wiley: 1998. Pp. 181. £29.95, \$55

Roy Gibson

For many years, almost anything connected with space was regarded in China as a state secret, and even seasoned veterans of the Chinese system had difficulty piecing together what little information emerged. Even since the opening up of China in the late 1970s, with foreign visitors and international meetings becoming not only accepted but actively encouraged, it has remained difficult to create an accurate picture of the Chinese space programme. The language barrier is an additional hurdle (although Brian Harvey has been admirably consistent in his transliterations), and one suspects that the Chinese have not been too unhappy at the confused picture presented to the West.

Harvey's book is a worthwhile acquisition for anyone interested in the development of space activities in China. It is packed with facts, nearly all of which are accurate, but regrettably it is rather short on analysis. Moreover, some of the facts included tend to be of the 'engine driver's name' variety, and there is also a disturbing imbalance in the depth of treatment in the various sections. As a source book it is to be highly recommended, but it is based on a rather limited and institutional view of space, so it fails to convey an adequate notion of the importance of space in China's overall development in the global context.

To assess the significance of space activities, it is nowadays necessary to look beyond the programme of the national or regional space agency. Important though they still are, the US space agency NASA and its European equivalent ESA now have little effect on the development of commercial space programmes. The myriad telecommunications, broadcasting, navigation and positioning satellite programmes have little if nothing to do with the space agencies (even though they nearly all started by reason of the space agencies' pioneering work). Even the development of launch capability has as much to do with potential profit as with providing a national launch means for government, civil and military satellites. All of this is increasingly true for China.

For the past two decades, Chinese space policy appeared to have multiple objectives. The original defence impulse remains as strong as ever, but the Chinese were also interested in harnessing the benefits of space technology as quickly as possible for the development of their country in general and



Rocket power: China's political ambitions exert a strong influence on the country's efforts in space.

their industry in particular. In so doing they wished to demonstrate to the region, indeed to the world, that they belonged to the élite group of 'space nations'. This was partly to win internal and external admiration, and partly as an effective advertisement for their growing commercial space ventures.

It is a pity that Harvey did not set out to examine these various topics instead of concentrating on the details of the events of the past 30 years. At least three aspects deserved attention. The first is the use of space capability as an instrument of foreign policy. Apart from the important 1996 International Astronautical Federation's Congress in Beijing — which, as Harvey points out, was a sort of coming-out party for the Chinese

space programme — China is hosting increasing numbers of United Nations space conferences, including some at ministerial level, that would previously have been held elsewhere in the region.

Of the many intergovernmental space agreements China has signed over the past 15–20 years, the one signed with Brazil in 1993 is a good example of political content outweighing the value in space terms. China has invested in impressive satellite integration facilities in Brazil (inaugurated by the Chinese president), as well as the joint development of the Sino-Brazilian Earth Resources Satellite. This investment cannot simply be explained as a means of moving forward China's own remote-sensing pro-

gramme. Skilled though they are in this field, the Brazilians had very little that the Chinese had not already developed or at least fully understood. It is difficult to see the agreement as anything other than a political investment, with perhaps the longer term goal of access to a launch site in the Southern Hemisphere.

Second, it would have been interesting for Harvey to have explored the telecommunications sector, the very slow break-up of the state monopoly, and the acceptance of foreign commercial investments to exploit satellite communication technology. This is probably the sector that will bring about the greatest visible change to life in China over the next few years. Ericsson, Motorola and their rivals worldwide all have a presence in China, and the once staid Ministry of Telecommunications is now on familiar terms with Iridium, Globestar and other global satellite networks.

Finally, I would have liked Harvey to have devoted a few pages to the opening up of space-related joint ventures, and to have examined the changes in Chinese internal organization and policy which these necessitated. Joint ventures range from the high-profile EuroSpace (a collaboration between Deutsche Aerospace and the China Aerospace Corporation) and Chinese government participation in the commercial satellite company Asia Pacific Mobile Telecommunications to more modest companies, such as the Canadian ComDev's joint venture with the Xi'an Institute of Space Radio Technology and the Shanghai Rockwell Collins company. Despite considerable bureaucracy at regional and national levels, the Chinese have been sufficiently flexible to put together business deals significantly more sophisticated than the usual agreements between space agencies.

By the same token, Harvey might have examined more closely the prospects for the commercial future of China's launcher industry. Despite setbacks, Chinese launcher services (now able to provide insurance and re-launch deals) are likely to acquire and retain about 10–15 per cent of the world market. The days of US-imposed launch limits are gone, thanks not least to US satellite-system investors eager to have reliable and cheaper launches, and the Chinese have a contract to launch some of the Iridium satellites.

Space will not escape the new government measures to spruce up the 370,000 state-run enterprises that this year's nomination of Zhu Rongji as prime minister will bring. A new super-ministry, the National Ministry of Defence Industries, will be created to take over the numerous factories working for the electronic, space and defence sectors. This may not seem to be a promising step towards reducing bureaucratic interference, but even local sceptics admit that it will

probably result in the delegation of more authority to the regions and challenge individual factories to make a profit or be absorbed. This will almost certainly lead to more joint ventures and more investment from overseas. Unfortunately, these recent events came too late for inclusion in Harvey's book.

All of this is some way removed from the historical detail of Harvey's book, but it is essential to an understanding of China's multi-pronged approach to space. As Harvey suggests in his final chapter, its future should be judged not on the successful achievement of manned space flight, but on progress in all of the space-related sectors that China has opened up. The pace in each sector will be different, reflecting the areas in which China sees the balance of advantage in any particular period. The basic objective — and an entirely realistic one — is for China to be a well-established and important space power early in the next century, with programmes in science, commercial applications, manned space flight... and defence. □

Roy Gibson is at *Résidence les Hespérides, 51 Allée Jean de Beins, F34000 Montpellier, France.*

Lost for words

Why Our Children Can't Read and What We Can Do About It

by Diane McGuinness
*Free Press/Penguin: 1997/1998. Pp. 396.
\$25 (hbk), £8.99 (pbk)*

Philip T. Smith

Diane McGuinness's two-year-old son used to tip over furniture "to see if it would break". There is no doubt from which parent he inherited this trait. In an impassioned book, McGuinness seeks to tip over the educational establishment that she holds responsible for the poor standards of literacy in the United States and Britain.

Few are exempt from criticism, from those responsible for government policy dating back to the nineteenth century that downgraded the status of teachers, to child-development experts who have no children of their own. The chief culprits are the advocates of teaching schemes ('look-and-say', 'phonics' and 'real books') that do not work (Kenneth Goodman, who said that reading is a "psycholinguistic guessing game", comes in for particular criticism), and also the vested interests in the educational establishment, particularly in government departments and teacher training colleges. Both groups are failing, according to McGuinness, because the methods they favour are based on inadequate research, and important research findings are ignored.

The crucial research can easily be summarized. Speech can be described as a sequence of segments called phonemes,

together with suprasegmental features responsible for such things as stress and intonation. Phonemes are difficult to define rigorously, but correspond to units that listeners perceive as having an invariant sound (so speakers of English perceive that the words 'pit', 'spit' and 'tip' each contain the phoneme /p/). An alphabet is a system for transcribing speech using letters to represent phonemes (so the letter 'p' often represents the phoneme /p/ in English).

We might think it should be easy to teach people to write with an alphabetic system: all they have to do is to listen to the phonemes in a word and transcribe them; reading should be the equally simple reverse process. But having sufficient awareness of the phonemes in a word to carry out this transcription process is a skill that does not come easily and needs to be carefully taught. Research in the past 30 years, principally by psychologists in Belgium, Britain, Sweden and the United States, has shown that fluency in manipulating phonemes goes hand in hand with mastering an alphabetic writing system: those who lack this 'phonemic awareness' will be severely limited in how far they can progress in learning to read, as they will lack the skills to handle unfamiliar words.

Deficits in reading ability, in McGuinness's view, are almost all due to inadequate teaching: "Everyone... can be taught to read unless they have such deficient mental and/or linguistic skills they can't carry on a normal conversation." An adequate teaching programme should consist of developing phonemic awareness, followed by a carefully phased introduction of the alphabetic system, with emphasis on letters being pictures of phonemes, and delaying exposure to some of the more complex idiosyncrasies of English spelling until the basic code is mastered.

She says that children who fail at reading do not have some deep-seated biological deficit (dyslexia) but have developed inadequate strategies as the result of poor teaching: children taught by the look-and-say or real books methods do not understand the alphabetic code because they are concentrating on units that are too large; children taught by 'phonics' stand a better chance of cracking the code, but the emphasis on sounding out letters ('letter→sound') can obscure the essential insight that an alphabet transcribes sounds ('sound→letter').

I applaud what McGuinness is trying to do, and consider, from a practical point of view, that she is correct. If a child (or adult) has reading difficulties, our first hypothesis, once gross sensory impairments have been discounted, is that reading instruction has been inadequate, and that a proper remedial scheme, based on developing phonemic awareness, is likely to succeed. But I think she is too hasty in discounting visual and biological contributions to reading difficulties: only three pages are devoted to vision and I

am surprised she makes no reference to the work of William Lovegrove, whose research suggests that subtle interactions between the magnocellular and parvocellular systems in vision can lead to reading difficulties.

Also, despite her references to the work of Paula Tallal, McGuinness does not see a weakness in the processing of rapidly changing stimuli (possibly in both the auditory and visual modalities) as a barrier to reading for all but a small minority of children. As in many areas of biology and psychology, nature–nurture interactions are probably more complex than we think.

The book is long, which may put off some parents; it lacks an index; and it is not always easy to match up claims in the text with references at the end of the book, which will put off some academics. But it is a very important book, and should be read by every member of the educational establishment. If McGuinness's ideas were fully implemented in schools, we would see substantial reductions in the number of children experiencing reading difficulty, with all its concomitant miseries. □

Philip T. Smith is in the Department of Psychology, University of Reading, Whiteknights, Reading RG6 2AA, UK.

Some like it hot

Life on the Edge: Amazing Creatures Thriving in Extreme Environments

by Michael Gross

Plenum: 1998. Pp. 200. \$25.95, £15.75

Andrew R. Cossins

Extremists of all kinds tend to challenge firmly held concepts and ideas. This is particularly true with the discovery in the past two or three decades of microorganisms that inhabit conditions previously thought to be totally incompatible with life. The first such organisms were isolated from hot springs in Yellowstone National Park by Thomas Brock. The surprise was not just that they live at temperatures of up to 95 °C, but that they cannot be cultured at lower temperatures; they have become thermal specialists *par excellence*, unable to cope with change.

This finding encouraged microbiologists to expand their travel budgets and explore a wide range of very uncomfortable environments, including boiling mud-holes, volcanoes, solfatara fields and even 'black smokers' on the ocean floor, all of which are home to microorganisms. Microbial communities have since been found in the extreme cold of the Antarctic, in the hypersaline Dead Sea in Israel, under the very high hydrostatic pressures of ocean trenches, and even in rock drillings from 1,500 metres beneath the US state of Washington. These astonishing discoveries extended the microbiological lexicon to include such terms as



Water colours

Jellyfish are "among the most beautiful animals of the sea" and are composed of 95% or more water, write David Wrobel and Claudia Mills in *Pacific Coast Pelagic Invertebrates* (Sea Challengers/Monterey Bay Aquarium, \$16.95 (pbk)).



extremophile, hyperthermophile and chemolithoautotroph. Michael Gross has written a highly readable account of these organisms and the excitement surrounding their discovery. His overview is masterfully broad, linking their discovery with a number of recent developments in molecular biology, biotechnology, pollution control, biogeochemistry and the origin of life.

Perhaps the most fascinating aspects of these microorganisms is how they thrive in conditions that would rapidly kill more familiar organisms. The proteins of hyperthermophiles must be particularly resistant to denaturing at high temperatures, although Gross shows that there is no simple or consistent set of structural features that account for this. He describes heat-shock proteins and their role in refolding heat-damaged proteins, although their specific

role in extremophiles is uncertain because they are found in virtually all living organisms, including those inhabiting polar habitats. Similarly, he discusses the antifreeze glycoproteins of polar fish, even though they have no role in microbial adaptation.

Yet Gross virtually ignores the issue of membrane adaptation through the modification of lipid structures, an important feature of adaptation to extremes of both temperature and pressure in a wide variety of animals, plants and microbes. Hyperthermophiles in particular possess very unusual membrane lipids, some of which appear to suppress the excessive molecular motion that high temperatures might otherwise cause.

Gross is on safer ground in showing that extremophiles are more than just a scientific curiosity. Their discovery has underpinned

the emerging recognition of a new branch of life, the Archaea; it has led to the discovery of submarine and cave-dwelling biotopes that are entirely independent of sunlight; and it has provided insights into the processes of planetary geochemical change. Moreover, extremophiles have potentially important implications for biotechnology and medicine, which are given extensive treatment. Hyperthermophiles are the source of that most valuable biotechnological product, thermostable DNA polymerase, and seem likely to deliver other innovative products and biotechnological processes.

Gross concludes by considering the possibilities for the existence of life in the Universe beyond Earth. The discovery of extremophiles has moved the goalposts in extending the range of conditions and environments considered compatible with life. Nevertheless, the technical problem remains of defining exactly how life, even of a simple microbial form, can be detected by remote probes; the most likely method is the discovery of unexpected gas compositions in planetary atmospheres.

Gross has a writing style that is both informal and entertaining. The text is enlivened by simple line diagrams and a series of vignettes describing a range of topics, including short anecdotal biographies of the principal researchers. The result is a tale of modern biology linking the improved understanding of microbial biodiversity to an impressive range of scientific, medical and commercial advances. It is written at a level that would suit non-specialists, even non-biologists, and therein lies its value: bringing a complex and developing tale of modern biology to a new audience.

Even uncomfortably extreme conditions of temperature, pressure and salinity can be optimal for some species. Perhaps the most challenging problems result not so much from survival under extreme but constant conditions, but from routine exposure to normal but more variable conditions. □

Andrew R. Cossins is in the School of Biological Science, University of Liverpool, Liverpool L69 3BX, UK.
e-mail: cossins@liverpool.ac.uk

Designer chemistry

Eloge du Mixte: Matériaux Nouveaux et Philosophie Ancienne

by Bernadette Bensaude-Vincent

Hachette Littératures: 1998. Pp. 358. FF130

Pierre Laszlo

From the sledgehammer to the skilled operator and the sophisticated robot, work in a scrapyard requires considerably more discernment than it did 30 years ago. Separating the various materials that make up a car — the steel, the aluminium, the platinum, the

polyurethane cushions, the vinyl sidings, the elastomers from the tyres, the polyester seat fabrics, the acrylic, the glass windows, the reinforcing glass fibres, and so on — has spawned new companies. Recycling has become a sizeable portion of the plastics industry.

This book addresses the topic of man-made materials, of composites in particular. Its timeliness stems from the advent during the past two decades of a materials science. Seen from physics, it embodied the hope of devising new and interesting problems from the study of 'soft matter', at a distance from traditional solid-state physics. Seen from chemistry, it looked like a new subdiscipline, parallel to molecular science, that would provide common ground for polymer scientists, solid-state chemists, experts in sol-gel processes and suchlike. Important discoveries then helped to boost and buttress materials science.

In 1971, Stephanie Kwolek invented aramides at DuPont de Nemours. The best known among these polyamides is Kevlar, a fibre with, weight for weight, three times the tensile strength of steel. The discovery of carbon nanotubes 20 years later provided materials hundreds of times more resistant than steel that can be turned into electrical wires at the nanometre scale. Yet another unpredicted advance was the discovery in 1986, far from the beaten path, of supraconducting ceramics by Johannes Georg Bednorz and Karl Alex Müller at IBM in Zürich.

The author of this book, Bernadette Bensaude-Vincent, a professor of philosophy and science historian, is well known as a specialist of eighteenth-century French chemistry, and of Antoine Lavoisier in particular. She was recently awarded the Dexter prize in science history by the American Chemical Society.

With this book on materials, she has now moved to an altogether different topic. The first quarter of the book goes back to Greek philosophy: wood is the archetypal material; composites have an ancient history too, going back to Plato's description of crafts in *Gorgias*, and manifest man's hubris, his daring to contrive artificial fabrics to rival natural products. The second part alludes to the history of the relevant concepts: man-made chemicals such as soda were first made in the eighteenth century; synthetic organic chemicals originated in the nineteenth century. The notion, which we deem modern, of science as the engine for technological innovation can actually be traced back to two Frenchmen, Bernard Palissy — a figure of the Renaissance and a 'Renaissance man' himself — and René Antoine Ferchault de Réaumur, an early theoretician of metallurgy in the eighteenth century.

The third section eulogizes our 'age of plastics', the design of reinforced composite materials and the underlying web of intricate

interactions between the component fibres and the associated interwoven surfaces. The final act questions, in the manner of an investigative reporter, the already mythical history of composite materials, seen as fall-out from rocket science and other military applications.

Like a lepidopterist, Bensaude-Vincent tends to pin down concepts and then refer to them for 'show-and-tell' stories. But intellectual history is more demanding: there is a desire to understand why the need arose for a given concept, to know how it became embodied in an object, to learn when and why it started to evolve, and to be reminded of its place in our present perception. Bensaude-Vincent is not always reliable in her assertions: alizarine had nothing to do with the synthesis of indigo, and she gets wrong the first name of Jorge Luis Borges.

Her book is easy to read; the style varies from the mundane, as when it smacks of a newspaper article or a report to stockholders, to a racy and supple French prose, as in her description of the art of the synthetic organic chemist, *Homo ludens* as she names him. But the contents are undermined by their conservatism.

She tries to examine the philosophical issues raised by mankind's new materials. Unfortunately, she does so with a rear-view mirror, showing the traditional fare — Plato for hors-d'oeuvre, Aristotle for the main course and Kant for dessert — instead of a view ahead, through a laminated windshield, at an innovative philosophy to go with the new materials in our life. □

Pierre Laszlo is in the Department of Chemistry, Ecole Polytechnique, 91128 Palaiseau, France.

New Journals

This year, *Nature's* annual new journals review supplement will appear in the issue of 10 September. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals must have first appeared during or after June 1996 and issued at least four separate numbers by the end of May 1998.
- Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are newsletters and publications of abstracts.
- Frequency of publication must be at least three times a year.
- The main language is English.
- Deadline for submission is 5 June.

Please send at least four different issues (the first, the most recent and any two others) of each eligible title, together with full details of subscription rates, to:

Peter Tallack, *Nature*, Porters South, Crinan Street, London N1 9XW, UK. Tel: +44 (0)171 843 4567. e-mail: p.tallack@nature.com

Role of a p53 polymorphism in the development of human papilloma-virus-associated cancer

Alan Storey*, Miranda Thomas†, Ann Kalita‡, Catherine Harwood*, Daniela Gardiol†, Fiamma Mantovani†, Judith Breuer§, Irene M. Leigh*, Greg Matlashewski‡|| & Lawrence Banks†||

* Imperial Cancer Research Fund, Skin Tumour Laboratory, 2 Newark Street, London E1 2AT, UK

† International Centre for Genetic Engineering and Biotechnology, Padriciano 99, I-34012 Trieste, Italy

‡ Institute of Parasitology and McGill Cancer Centre, McGill University, Macdonald Campus, 21,111 Lakeshore Road, Ste Anne d Bellevue, Québec H9X 3V9, Canada

§ Department of Medical Microbiology, St Bartholomew's and The Royal London Hospital School of Medicine and Dentistry, Queen Mary and Westfield College, Ashfield Street, London E1 1BB, UK

|| These authors contributed equally to this work

The E6 oncoprotein derived from tumour-associated human papillomaviruses (HPVs) binds to and induces the degradation of the cellular tumour-suppressor protein p53. A common polymorphism that occurs in the p53 amino-acid sequence results in the presence of either a proline or an arginine at position 72. The effect of this polymorphism on the susceptibility of p53 to E6-mediated degradation has been investigated and the arginine form of p53 was found to be significantly more susceptible than the proline form. Moreover, allelic analysis of patients with HPV-associated tumours revealed a striking overrepresentation of homozygous arginine-72 p53 compared with the normal population, which indicated that individuals homozygous for arginine 72 are about seven times more susceptible to HPV-associated tumorigenesis than heterozygotes. The arginine-encoding allele therefore represents a significant risk factor in the development of HPV-associated cancers.

The development of cervical carcinoma correlates closely with the presence of certain human papillomavirus types, such as HPV-16 and HPV-18. In contrast, infection with other HPV types, such as HPV-6 and HPV-11, correlates mainly with the development of benign lesions¹. HPV-16 and HPV-18 encode two major oncoproteins, E6 and E7. The E7 protein binds to and inactivates the cellular tumour-suppressor protein Rb (ref. 2); the E6 protein binds to the cellular tumour-suppressor protein p53 and directs its degradation through the ubiquitin pathway^{3,4}. Although inactivation of these two tumour-suppressor proteins by HPV is probably important for tumour development, no genetic factors have been conclusively identified that might predispose an infected individual to develop cervical carcinoma. p53 protein is mutated in many human tumours but is usually wild-type in early cervical tumours⁵⁻⁷, indicating that inactivation of p53 by E6 is analogous to an inactivating mutation. Many mutant p53 proteins are not susceptible to E6-mediated degradation⁸⁻¹⁰, suggesting that the interaction between these two proteins can be easily disrupted.

When wild-type p53 complementary DNAs were first cloned from normal human cells, a sequence polymorphism was found in the general population that results in either a proline (p53Pro) or arginine (p53Arg) at residue 72 and affects the migration of the protein on SDS-PAGE¹¹. We therefore investigated whether this polymorphism could affect p53 degradation mediated by E6 and if this might be reflected in the p53 genotype in HPV-associated tumours. Screening of a wide variety of tumours has not yet linked either allele with a susceptibility to any particular tumour, although these alleles may be markers in linkage disequilibrium with cancer-susceptibility genes, distinct from p53, in lung cancer¹². Furthermore, the allelic variants seem to be functionally equivalent¹³.

Allelic susceptibility to degradation *in vitro*

To test whether the E6 protein from HPV-18 preferentially recognizes either of the two forms of p53, we studied their degradation *in vitro*. The HPV-18 E6 and p53 proteins were translated *in vitro* and

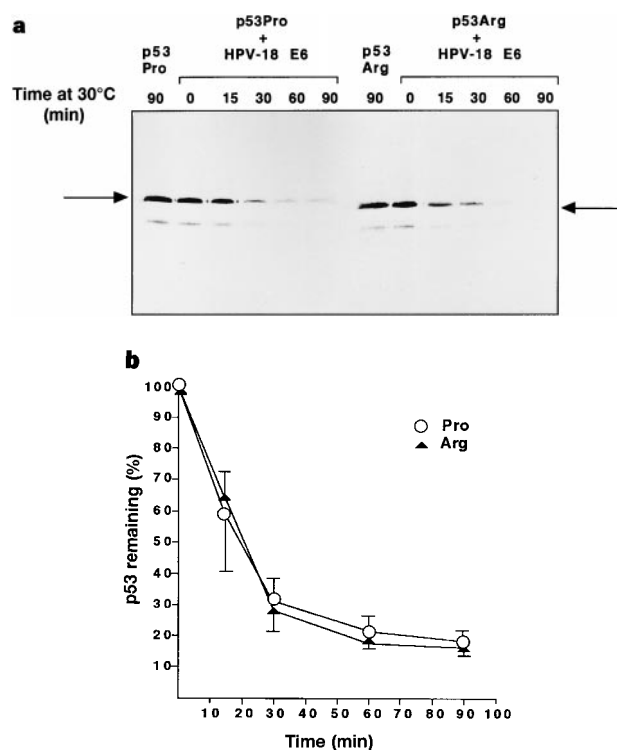


Figure 1 Comparison of the rate of HPV-18 E6 mediated degradation of p53Pro and p53Arg *in vitro*. Equal amounts of *in vitro* translated p53Pro and p53Arg were incubated with HPV-18 E6 at 30°C for the times indicated. **a**, The remaining p53 was determined by immunoprecipitation using pAb1801 (ref. 44), followed by PAGE and autoradiography. **b**, Quantification by Phosphor Imager of the remaining p53 protein after incubation with HPV-18 E6 for the times indicated. Numbers shown represent the mean from five separate experiments with standard deviations.

degradation of p53 was induced by addition of HPV-18 E6; reactions were terminated after the times indicated (Fig. 1a) and remaining p53 was quantified by immunoprecipitation (Fig. 1b). Although p53Pro seems to be slightly more resistant to E6-mediated degradation than p53Arg (Fig. 1a), time-course assays (Fig. 1b) revealed only minor differences in the susceptibility of the two p53 proteins to E6-mediated degradation *in vitro*.

Allelic susceptibility to degradation *in vivo*

p53 mutants may differ in their susceptibility to E6-mediated degradation, depending upon whether they are assayed *in vitro* or *in vivo*¹⁴. We therefore transfected p53-null Saos-2 cells with plasmids expressing either p53Pro or p53Arg, together with increasing amounts of an HPV-18 E6-expressing plasmid. After 24 hours, we extracted the cells and analysed the residual p53 protein by western blotting. Figure 2 shows that increasing amounts of HPV-

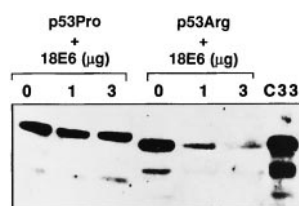


Figure 2 Comparison of HPV-18 E6 induced degradation of p53Pro and p53Arg *in vivo*. Saos-2 cells were transfected with 3 μg of either pCDNA-p53Pro or pCDNA-p53Arg plus increasing amounts of pCDNA-18E6 or control plasmid pCDNA-3 (lane 0), as indicated. After 24 h, cells were extracted and p53 protein detected by western blot analysis with a pool of anti-p53 monoclonal antibodies. C33 cells represent a positive control for p53 expression.

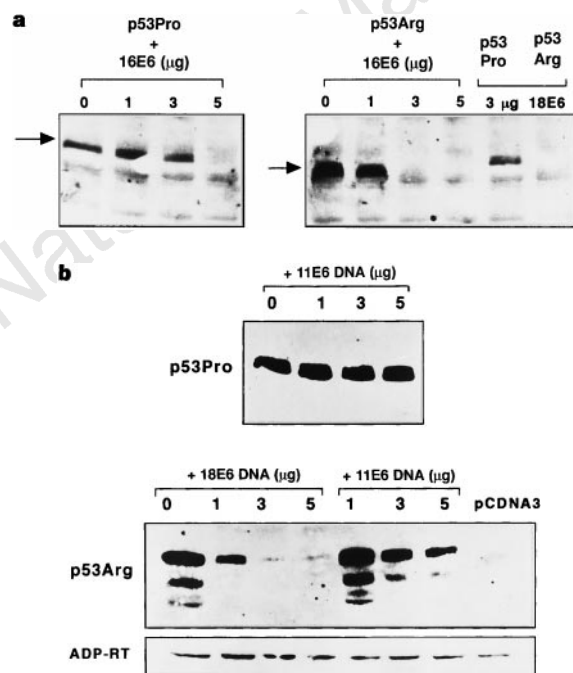


Figure 3 HPV-16 and HPV-11 E6 preferentially target p53Arg over p53Pro for degradation *in vivo*. Saos-2 cells were transfected with pCDNA-p53Pro or pCDNA-p53Arg, as indicated, together with increasing amounts of **a**, pCDNA-16E6, or **b**, pCDNA-11E6, or control plasmid pCDNA-3, as indicated. Residual p53 protein was detected as for Fig. 2. pCDNA-18E6 was included for comparison. The lower panel shows the same blot reprobed with an antibody directed against ADP-RT to demonstrate equal protein loading.

18 E6 cause a dramatic decrease in the remaining p53Arg, consistent with our *in vitro* results. In contrast, there was no significant degradation of p53Pro under identical conditions. These results indicate that the proline/arginine polymorphism at position 72 of wild-type p53 affects the proteins' susceptibility to HPV-18 E6-induced degradation *in vivo*, with p53Arg being more susceptible than p53Pro.

To test whether this susceptibility of p53 to E6-induced degradation was peculiar to E6 from HPV-18, we assayed degradation *in vivo* with increasing amounts of E6 from HPV-16 and HPV-11. We found that p53Pro was more resistant to HPV-16 E6 than the p53Arg variant and that degradation of p53Pro was only significant at high input levels of E6, whereas much less HPV-16 E6 was needed for complete degradation of p53Arg (Fig. 3a). Previous *in vitro* studies indicated that HPV-11 E6 could not target p53 for ubiquitin-mediated degradation³. Our results *in vivo* showing that HPV-11 E6 could not mediate degradation of p53Pro support this view (Fig. 3b). However, HPV-11 E6 does induce degradation of p53Arg *in vivo*, although not as efficiently as the oncogenic E6 from HPV-16 and HPV-18. These results show that the proline/arginine polymorphism in wild-type p53 protein at position 72 affects the susceptibility of p53 to E6-mediated degradation, regardless of whether infection is by a low-risk or high-risk HPV type.

Upon labelling with ubiquitin, proteins are rapidly degraded by

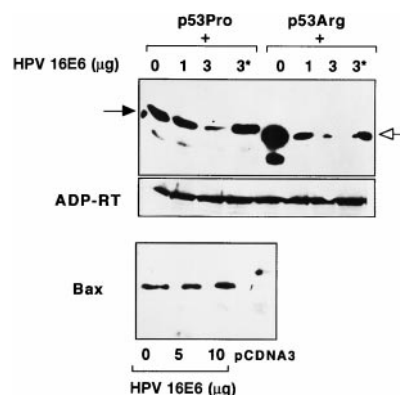


Figure 4 HPV-16 E6-induced degradation of p53Pro and p53Arg is mediated by the proteasome. Saos-2 cells were transfected with either pCDNA-p53Pro or pCDNA-p53Arg, as indicated, together with increasing amounts of pCDNA-16E6. Tracks (asterisks) included treatment with the proteasome inhibitor LLnL for 2 h before protein extraction. Residual p53 protein was detected as for Fig. 2. The lower panel is the same blot reprobed with an antibody directed against ADP-RT to demonstrate equal loading. Saos-2 cells were also transfected in parallel with pCDNA-HA-Bax plus increasing amounts of pCDNA-16E6 as indicated (lower panel), and western blot analysis was performed with an anti-HA antibody to detect Bax protein expression.

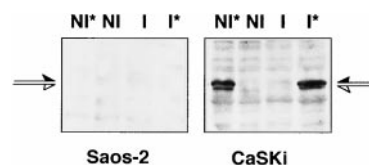


Figure 5 Comparison of the susceptibility of p53 to E6-mediated degradation in a cell line containing endogenous p53Pro and p53Arg alleles and HPV-16 E6. The levels of p53Pro and p53Arg protein in CaSKI cells (right) were assessed by western blotting. Cells were either untreated (NI), or treated for 18 h with mitomycin C (I). The proteasome inhibitor, lactacystin (*), was added 2 h before collecting cells. Solid arrow, p53Pro; open arrow, p53Arg. The pattern of lower *M_r* background bands demonstrates equal loading. As a control, p53-null Saos-2 cells were treated in parallel (left).

the proteasome complex^{15,16}. To verify that the E6-induced degradation of p53 occurs by the ubiquitin pathway, we repeated the degradation assay with HPV-16 E6 and cells incubated with the peptide aldehyde proteasome inhibitor *N*-acetyl-Leu-Leu-norleucinal (LLnL)¹⁷, added 22 h post-transfection. At 24 h, we extracted the proteins and quantified p53 by western blot analysis. As an additional control, a plasmid expressing Bax protein was co-transfected with the HPV-16 E6-expressing plasmid (Fig. 4). Again, p53Arg was more susceptible to E6-induced degradation than p53Pro; the parallel transfection of Bax plus HPV-16 E6 did not alter the amount of Bax protein in the assay, indicating that degradation is specific to p53. Addition of proteasome inhibitor protected p53 from E6-induced degradation, confirming that p53 was being degraded by the proteasome.

These effects on p53 were obtained in cells that were expressing either p53Pro or p53Arg. We also tested the effects of E6 on a cell that was heterozygous for the polymorphism. As p53 normally exists as a tetramer^{18,19}, we assumed that p53Pro and p53Arg form heteromers with equal efficiency and that the p53 pool consists of a mix of homo- and heterotetramers. To investigate the effects of E6 on p53 heteromers, we used a cell line that stably expresses heterozygous p53 and HPV-16 E6, namely CasKi, which is derived from a human cervical tumour, expresses HPV-16 E6 (ref. 20), and contains both polymorphic forms of p53 (as determined by polymerase chain reaction after reverse transcription (RT-PCR) and DNA sequencing; data not shown). p53 protein was induced by mitomycin C treatment²¹ in the presence or absence of proteasome inhibitor and quantified by western blotting (Fig. 5). Without induction, only a small amount of p53Pro was detected; after mitomycin C treatment both forms of p53 were weakly induced. However, following addition of the proteasome inhibitor

lactacystin²², a doublet of p53 became visible, with more of the upper band (p53Pro) being present than the lower (p53Arg) band. We conclude that in a heterozygous situation p53Pro is more resistant than p53Arg to E6-mediated degradation. It remains to be established whether this is the case for other cervical-derived cell lines or for cells immortalized in culture with different HPV types.

Genotypes prevalent in cervical tumours

As p53 is crucial for host defence against tumours the p53Pro/p53Arg polymorphism should influence the outcome of HPV infection. In tumours that are not HPV-associated, no significant correlation has been found with the presence or absence of a particular polymorphic form of p53 (refs 12, 23–26). In contrast, our results indicate that in HPV-associated tumorigenesis the polymorphism at position 72 in wild-type p53 may present a significant risk factor.

We tested this idea by analysing the p53 genotype of cervical tumours using a PCR-based assay that specifically detects either the p53Pro or p53Arg allele; Fig. 6a shows the location of the primers used to identify the alleles and the position of the polymorphism with respect to the p53 functional domains. The frequency of the two alleles in a series of cervical cancer biopsies was compared with the frequency in the control group (Table 1). As p53 allele frequencies vary according to ethnic group, we analysed controls and cervical cancer patients from the same ethnic background. The allelic frequencies in our control samples were similar to those previously reported^{25–27}. We found a marked difference in the frequency of Pro/Arg alleles between cervical cancer biopsies and normal controls ($\chi^2 = 10.5$, $P < 0.0005$; 95% confidence interval, 2.19–22.59). The p53Arg allele alone was detected in 76.7% of tumours, compared with 36.6% of controls; the Pro/Arg genotype

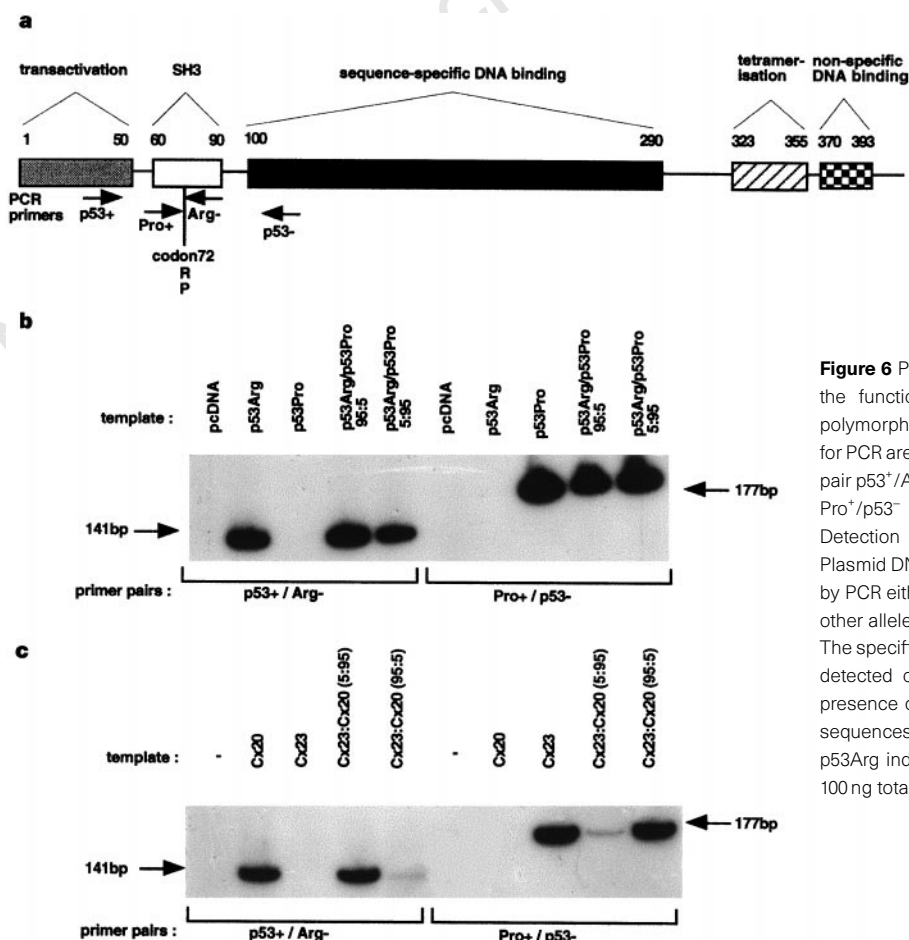


Figure 6 PCR detection of p53 alleles. **a**, The *p53* gene, showing the functional domains of the protein. The location of the polymorphism at codon 72 and the positions of the primers used for PCR are indicated. For amplification of the Arg allele, the primer pair p53⁺/Arg⁻ gives a PCR product of 141 bp; the primer pairing of Pro⁺/p53⁻ amplifies a 177-bp fragment of the proline allele. **b**, Detection of p53Pro and p53Arg variants from plasmid DNA. Plasmid DNA containing either the Pro or Arg allele was amplified by PCR either singly or in the presence of plasmid containing the other allele, using primers that specifically detect only one allele. The specific products of the Arg (141 bp) or Pro (177 bp) allele were detected only from plasmid containing that allele even in the presence of excess competitor template. **c**, Amplification of p53 sequences from genomic DNA from homozygous p53Pro or p53Arg individuals. Mixing of templates was done as in **b** using 100 ng total genomic DNA as the larger input.

Table 1 Frequency of Arg and Pro p53 alleles in normal and tumour tissue

	Pro	Arg	Pro/Arg
Normal tissue, N = 41	2 (0.05)	15 (0.37)	24 (0.58)
Cervical tumours, N = 30	2 (0.07)	23 (0.76)	5 (0.17)

The relative frequency of each allele is shown in parentheses. HPV typing of the cancers showed that HPV16 was the most prevalent (68%), followed by HPV18 (22%) and HPV45 (2%). The primer pairings used to detect the p53 Pro and Arg alleles are shown in Fig. 6a. Mucosal HPV types were detected using the primer pairs MY09/MY11 (ref. 45) and GP5/GP6 (ref. 46).

was found in 16.7% of tumours compared with 58.5% of controls. On the basis of these findings, with other risk factors being equal, an individual homozygous for p53Arg would be about 7 times more likely to develop HPV-associated cervical cancer than a Pro/Arg heterozygote. We detected no significant difference in the frequency of the homozygous p53Pro allele, although the numbers in each group were small.

p53Arg genotype is not due to loss of heterozygosity

Biopsies of tumours are invariably contaminated with stromal tissue or inflammatory cells, so we mixed p53Pro and p53Arg DNA templates in a single reaction to determine the level of contamination by non-tumour cells. First, we did PCR using plasmid DNA containing p53Arg and p53Pro sequences mixed at a ratio of 95:5 or 5:95, which showed that the product derived from the lower-copy-number allele was amplified to the same degree as that from the higher-copy-number allele (Fig. 6b), indicating that amplification of one allele was not affected by the presence of the other and that the different alleles were amplified to the same extent by the two primer pairs. To assess stromal contamination of biopsy material we used two methods. First, genomic DNA samples homozygous for either p53Pro or p53Arg were mixed at ratios of 95:5 and 5:95, when the minor DNA component was amplified much less than the major DNA component (Fig. 6c). Intensity of PCR products derived from heterozygous individuals was usually comparable and never varied more than twofold. Second, we analysed microdissected normal tissue and tumour biopsy material and found that the genotypes of both were largely identical, which agreed with the findings from quantitative PCR of DNA from frozen tissue that had not been microdissected. As the p53Pro allele was lost from tumour tissue in only 2/24 samples, the rate of loss of heterozygosity (LOH) at this locus was low, indicating that this is not an important mechanism for overrepresentation of p53Arg genotypes in cervical tumours; instead, it suggests that these tumours arose in individuals who were predisposed by having a homozygous p53 genotype.

The difference in the frequency of the two p53 alleles in cervical tumour and normal control samples indicates that the p53Arg allele, in the absence of the p53Pro allele, confers a susceptibility to cervical tumours in individuals harbouring high-risk HPV types. Such a difference in polymorphic frequency has not been found in other types of tumour, although a difference in allelic frequency at p53 codon 72 has been noted in non-smoking lung cancer patients with a higher incidence of Arg/Arg²⁷; this is set against other reports of an increased risk of between 1.5 and 1.9 for lung cancer patients associated with Pro/Pro^{28,29}. The greater susceptibility of the p53Arg protein to E6-induced degradation correlates with our clinical findings, suggesting that this degradation of p53 is important in the development of cervical tumours.

Different genotypes in HPV skin cancers

In addition to their role in cervical cancer, HPVs have also been implicated in the development of squamous cell carcinoma in renal transplant recipients (RTRs)³⁰ at cutaneous sites exposed to ultraviolet radiation; these tumours colocalize with and may originate from viral warts^{31–34}. In these immunosuppressed individuals, a broad range of HPV types occur in such lesions, unlike cervical

Table 2 Arg and Pro alleles of p53 in SCCs of renal transplant recipients

Patient	SCC code	HPV type(s)	Arg/Pro status of tumour (R/P)
C.W.	2H	27	R
	7S	10	R
	7U	28	RP
C.R.	2L	HPVRTR × 5*	R
A.Q.	3D	3, 20, 23	RP
D.E.	3U	10, HPVRTR × 5*	R
	3V	HPVRTR × 5*	R
	3X	HPVRTR × 1*	R
	7L	10, 23	R
	7M	5	R
D.L.	7Q	10, 23	R
	3Y	1, 10	R
	3Z	1, 27	R
B.M.	4A	27	R
	4B	27, 36	R
	4E	19, 27	R
	7F	37	R
	7G	1, 14D, 19	R
B.W.	7T	19, 66	R
	4N	5, 27, 77	R
	4P	77	R
D.P.	4T	27	R
	5C	24, HPVVS42L1*	R
R.A.	5D	24, 38, 77	R
	F	5, 10, 27	R
R.W.	7B	14D, 77, HPVVS92L1*	R
	7C	14D	R
	7D	10, 14D, 24	R
L.S.	7J	38-rel	RP
	7K	38-rel, HPVVS92L1*	RP
	7R	38-rel	R
A.B.	5J	23, 77	R

SCC, squamous cell carcinoma. Note that, unlike cervical tumours which contain only one HPV type, individual cutaneous lesions derived from different sites and at different times from the same patient often contain more than one HPV type. Cutaneous HPV types were detected using the outer primer pair HVP2/B5 (refs 35, 47) and 3 pairs of nested internal primers designed to detect the HPV groups A2 (ref. 48) (CN3F: 5'-AACTCTAAYATWGCAC ATG/CN3R: 5'-CAVGTRCSYTGCAATATC), A4 (CN2F: 5'-GGGGATATGGTTGAACAG GT/CN2R: 5'-CAGAGGACACCATAGAGCCA) and E (CN1F: 5'-AATARGTTWGATGATGCGW AA/CN1R: 5'-AKRTARTCWGGATATTGCA). The cutaneous HPV group B2 was detected using a single round primer pair (C4F: 5'-GGAGATACAGAAAATCCT/C4R: 5'-SHATCTCCAT AGATATCTTT). EV-associated HPV types of group B1 were detected with the nested primer pairs CP62/69 and CP65/CP68 (ref. 36) and 3 additional internal nested primer pairs designed to detect the major EV-HPV clusters (EN1F: 5'-TATTTCCCWACHGTHAGTGG CTC/EN1R: 5'-TCATAYTCYCTACATGTCT; ENF: 5'-CTGTCAGTGGCTCATTGGT/EN2R: 5'-CATWGCATTAATTGAGCTA; and EN3F: 5'-ATGKCAATGATGTHATGG/EN3R: 5'-TGRTT RYCCAYAAAATGCCATT), where W is A or T, S is C or G, H is A, C or T, R is A or G, K is G or T, Y is C or T, and V is A, C or G.

* Locus name for an HPV sequence available in the GenBank database for which the full genome sequence has not yet been published (accession numbers are given in parentheses): HPVRTR × 1 (L38918), HPV group⁴⁸ B1; HPVRTR × 5 (L38922), HPV group B1; HPVVS42L1 (X79943), HPV group B1; HPVVS92L1 (X79943), HPV group B1.

tumours which often contain only one HPV type. In addition to previously defined HPV types, several uncharacterized HPV sequences have been found in biopsies from RTRs^{35–37}. We therefore extended our polymorphism studies to include tumours derived from RTRs (Table 2). These results again show that the Arg allele confers a strong susceptibility to tumour development, with a striking excess of patients having only Arg (75%), rather than Arg/Pro (25%) or Pro (0%) alleles. However, the relationship between p53 status and the presence of HPV in these cutaneous lesions has not yet been established.

Discussion

We have demonstrated a difference in the susceptibility of two polymorphic forms of wild-type p53 to E6-mediated degradation *in vivo*: E6 from HPV-16 and HPV-18 is more effective at degrading p53Arg than p53Pro *in vivo*; HPV-11 E6 is less active towards p53Arg and inactive with p53Pro. We cannot yet explain the

different susceptibility of the two p53 forms *in vivo*, but the difference between degradation *in vitro* and *in vivo* is not without precedent¹⁴. The marked change in structure conferred on p53 by the Pro/Arg polymorphism, as shown by migration on SDS-PAGE, could alter the binding of E6, E6AP or other components of the ubiquitin pathway to p53 *in vivo*. When the sequence between residues 66 and 100 of p53 is lost, it is no longer susceptible to E6-induced degradation, indicating that the sequences adjacent to the polymorphism are needed for E6 to target p53 for degradation³⁸. This polymorphism lies in a region of p53 that is involved in the induction of apoptosis^{39,40} and which shares marked similarity to an SH3(SRC-homology-3)-binding domain⁴¹: it comprises five PxxP motifs, one of which is lost in the replacement of proline with arginine. We are currently investigating the effect of this polymorphism on the biological activity of wild-type p53. This is of interest because a germline mutation of p53 at residue 82 has been identified in a breast cancer pedigree⁴².

Our results indicate that p53Arg may represent a risk factor for HPV-associated tumorigenesis, as shown by p53 genotype analysis on two different HPV-associated tumours. In cervical tumours, whose aetiology is linked with certain types of HPV, there was a marked overrepresentation of homozygous p53Arg compared with heterozygous or homozygous p53Pro alleles, making individuals homozygous for p53Arg 7 times more likely to develop HPV-associated cervical cancer than individuals having one or more p53Pro alleles. Renal transplant recipients who are p53Arg homozygotes are also at increased risk of developing tumours (squamous cell carcinomas). Epidemiological surveys should be undertaken in larger populations and in different geographical regions to determine the role of this p53 polymorphism in HPV-associated lesions and whether the p53 genotype affects persistence of HPV infection and the development of premalignant lesions. □

Methods

Plasmids. Cloning and sequencing of wild-type p53Pro and p53Arg cDNAs has been described¹¹. For *in vitro* expression, p53 and HPV-18 E6 sequences were cloned into plasmid pSP64. For *in vivo* expression, p53 and HPV-11 E6, HPV-16 E6, HPV-18 E6 and HA-tagged Bax⁴³ sequences were cloned into plasmid pCDNA-3.

In vitro degradation assays. p53Pro, p53Arg and HPV-18 E6 were translated *in vitro* using the TNT system (Promega). Equal quantities of p53Pro or p53Arg were incubated with HPV-18 E6 at 30 °C, as indicated in the text. Residual p53 protein was then detected by immunoprecipitation with pAb1801 (ref. 44), followed by SDS-PAGE and quantification on a Phosphor Imager.

In vivo degradation assays. p53-null Saos-2 cells were transfected by calcium phosphate precipitation with 3 µg of p53 expression plasmids plus increasing quantities of HPV E6 expression plasmids, as indicated in the text. After 24 h, cells were extracted in a solution of 50 mM HEPES, pH 7.0, 250 mM NaCl, 0.1% N-P40 and 1% aprotinin. Protein concentrations were determined using the Bio Rad protein assay and equal amounts were analysed by western blotting. p53 protein was detected using a pool of anti-p53 monoclonal antibodies: pAb1801, 1802 and 1803 (ref. 44), and developed using the Amersham ECL system.

Cells were incubated with N-acetyl-Leu-Leu-norleucinal (LLNL, 50 µM), mitomycin C (10 µg ml⁻¹) and lactacystin (5 µM), as indicated in the text.

Transfection efficiencies were monitored by cotransfecting a lacZ-expressing plasmid (pCH110, Pharmacia) on parallel plates and staining for LacZ expression. C33 cells represent a positive control for p53 expression. To verify that protein loading was equivalent, western blots were reprobed with an anti-ADP-ribosyl transferase (ADP-RT) antibody.

Tissue samples. Tumour specimens included both invasive cervical and cutaneous squamous cell carcinomas. All lesions were histologically confirmed. Frozen specimens were collected at the time of biopsy or surgical excision, snap-frozen and stored at -70 °C. Samples for the control group consisted of DNA extracted from whole blood collected from healthy volunteers.

DNA extraction. DNA from frozen tissue was prepared by digestion with proteinase K and phenol-chloroform extraction, and from paraffin-embedded

tissue following deparaffinization. Identification of the relevant tissue for microdissection was aided using an adjacent stained section. DNA was extracted from whole blood from control individuals using a Nucleon DNA-extraction kit (Scotlab).

HPV detection and typing. All tumour specimens were tested for the presence and type of HPV DNA by using a degenerate PCR-based method with a panel of oligonucleotide primers located within the highly conserved L1 open reading frame of the HPV genome. Amplified PCR products were then sequenced on a Perkin-Elmer 377 ABI Prism automated sequencer.

PCR amplification of p53 polymorphic sequences. p53 Pro sequences were detected by PCR using the primer pair p53Pro+/p53- (p53Pro+: 5'GCC AGAGGCTGCTCCCC; p53-: 5'CGTGCAAGTCACAGACTT), and p53Arg by the primer pair: p53+/Arg- (p53+: 5'TCCCCCTTGCCGTCCCAA; p53Arg-: 5'CTGGTGCAGGGGCCACGC). Primer pairs (2 µg each) were mixed and end-labelled with ³²P by polynucleotide kinase (Promega) in a total volume of 50 µl containing 50 µCi [α-³²P] (Amersham); 1 µl labelled primer mix was used per PCR reaction. PCR (25 cycles) was done in 50 µl using 100 ng DNA template, 0.2 units Red Hot polymerase (Advanced Biotechnologies) in reaction buffer IV (supplied by the manufacturer) at 1.5 mM Mg²⁺. 10 µl reaction product was fractionated on an 8% polyacrylamide gel in 1× TBE buffer, dried, and either exposed to X-ray film or quantified using a Phosphor Imager. The χ² test was used to examine associations between genotype and the presence of cancer; odds ratios were also calculated.

Received 29 January; accepted 7 April 1998.

- zur Hausen, H. & Schneider, A. in *The Papovaviridae* (eds Salzman, N. P. & Howley, P. M.) 245–263 (Plenum, New York, 1987).
- Dyson, N., Howley, P. M., Münger, K. & Harlow, E. The human papillomavirus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science* **243**, 934–936 (1989).
- Scheffner, M., Werness, B. A., Huibregste, J. M., Levine, A. J. & Howley, P. M. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* **63**, 1129–1136 (1990).
- Werness, B. A., Levine, A. J. & Howley, P. M. Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science* **248**, 76–79 (1990).
- Crook, T., Wrede, D. & Vousden, K. H. p53 point mutation in HPV negative human cervical carcinoma cell lines. *Oncogene* **6**, 873–875 (1991).
- Scheffner, M., Münger, K., Byrne, J. C. & Howley, P. M. The state of the p53 and retinoblastoma genes in human cervical carcinoma cell lines. *Proc. Natl Acad. Sci. USA* **88**, 5523–5527 (1991).
- Vogelstein, B. & Kinzler, K. W. p53 function and dysfunction. *Cell* **70**, 523–526 (1992).
- Medcalf, E. A. & Milner, J. Targeting and degradation of p53 by E6 of human papillomavirus type 16 is preferential for the 1620+ p53 conformation. *Oncogene* **8**, 2847–2851 (1993).
- Marston, N. J., Crook, T. & Vousden, K. H. Interaction of p53 with MDM2 is independent of E6 and does not mediate wild type transformation suppressor function. *Oncogene* **9**, 2707–2716 (1994).
- Thomas, M., Massimi, P., Jenkins, J. & Banks, L. HPV-18 E6 mediated inhibition of p53 binding activity is independent of E6 induced degradation. *Oncogene* **10**, 261–268 (1995).
- Matlashewski, G. J. *et al.* Primary structure polymorphism at amino acid residue 72 of human p53. *Mol. Cell. Biol.* **7**, 961–963 (1987).
- Birgander, R. *et al.* p53 polymorphisms and haplotypes in lung cancer. *Carcinogenesis* **16**, 2233–2236 (1995).
- Moreau, F. & Matlashewski, G. Molecular analysis of different allelic variants of wild-type human p53. *Biochem. Cell.* **70**, 1014–1019 (1992).
- Crook, T., Ludwig, R. L., Marston, N., Willkomm, D. & Vousden, K. H. Sensitivity of p53 lysine mutants to ubiquitin-directed degradation targeted by human papillomavirus E6. *Virology* **217**, 285–292 (1996).
- Goldberg, A. The mechanism and function of ATP dependent proteases in bacterial and animal cells. *Eur. J. Biochem.* **203**, 285–292 (1992).
- Hershko, D. & Ciechanover, A. The ubiquitin system for protein degradation. *Annu. Rev. Biochem.* **61**, 761–807 (1992).
- Rock, K. *et al.* Inhibitors of the proteasome block degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell* **78**, 761–771 (1994).
- Milner, J. & Medcalf, E. A. Cotranslation of activated mutant p53 with wild type drives the wild type p53 into the mutant conformation. *Cell* **65**, 765–774 (1991).
- Sakamoto, H., Lewis, M. S., Kodama, H., Appella, E. & Sakaguchi, K. Specific sequences from the carboxy terminus of human p53 gene product form anti-parallel tetramers in solution. *Proc. Natl Acad. Sci. USA* **91**, 8974–8978 (1994).
- Androphy, E. J., Hubbert, N. L., Schiller, J. T. & Lowy, D. R. Identification of the HPV16 E6 protein from transformed mouse cells and human cervical carcinoma cell lines. *EMBO J.* **6**, 989–992 (1987).
- Butz, K. *et al.* Functional p53 protein in human papillomavirus-positive cancer cells. *Oncogene* **10**, 927–936 (1995).
- Fenteany, G. *et al.* Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. *Science* **268**, 726–731 (1995).
- Zhang, W., Hu, G. & Deisseroth, A. Polymorphism at codon 72 of the p53 gene in human acute myelogenous leukemia. *Gene* **117**, 271–275 (1992).
- Vogan, K. *et al.* Absence of p53 gene mutations in primary neuroblastomas. *Cancer Res.* **53**, 5269–5273 (1993).
- Beckman, G. *et al.* Is p53 polymorphism maintained by natural selection? *Hum. Hered.* **44**, 266–270 (1994).
- Weston, A. *et al.* Determination of the allelic frequencies of an L-myc and a p53 polymorphism in human lung cancer. *Carcinogenesis* **14**, 583–587 (1994).
- Murata, M. *et al.* Analysis of a germ line polymorphism of the p53 gene in lung cancer patients: discrete results with smoking history. *Carcinogenesis* **17**, 261–264 (1996).
- Kawajiri, K., Nakachi, K., Imai, K., Watanabe, J. & Hayashi, S. Germ line polymorphisms of p53 and CYP1A1 genes involved in human lung cancer. *Carcinogenesis* **14**, 1085–1089 (1993).

29. Jin, X. *et al.* Higher lung cancer risk for young African-Americans with the Pro/Pro p53 genotype. *Carcinogenesis* **16**, 2205–2208 (1995).
30. Purdie, K. *et al.* Malignant transformation of cutaneous lesions in renal allograft patients: a role for human papillomavirus. *Cancer Res.* **53**, 5328–5333 (1993).
31. Boyle, J., Briggs, J. D., Mackie, R. M., Junior, B. J. R. & Aitchison, T. C. Cancer, warts and sunshine in renal transplant patients. *Cancer Res.* **53**, 5328–5333 (1993).
31. Blessing, K. *et al.* Histopathology of skin lesions in renal allograft recipients—an assessment of viral features and dysplasia. *Histopathology* **14**, 129–139 (1989).
33. Benton, E. C. *et al.* in *Immunodeficiency and the Skin* (eds Fritsch, P., Schuler, G. & Hintner, H.) 168–177 (Karger, Basel, 1989).
34. Glover, M. T., Proby, C. M. & Leigh, I. M. Skin cancer in renal transplant patients. *Cancer Bull.* **45**, 220–224 (1993).
35. Shamanin, V. *et al.* Specific types of human papillomavirus found in benign proliferations and carcinomas of the skin in immunosuppressed patients. *Cancer Res.* **54**, 4610–4613 (1994).
36. Berkhout, R. J. M. *et al.* Detection and typing of epidermodysplasia verruciformis-associated human papillomavirus types in cutaneous cancers from renal transplant recipients. *J. Clin. Microbiol.* **33**, 690–695 (1995).
37. Shamanin, V. *et al.* Human papillomavirus infections in nonmelanoma skin cancer patients from renal transplant recipients and nonimmunosuppressed patients. *J. Natl. Canc. Inst.* **88**, 802–811 (1996).
38. Li, X. & Coffino, P. High-risk human papillomavirus E6 has two distinct binding sites within p53, of which only one determines degradation. *J. Virol.* **70**, 4509–4516 (1996).
39. Walker, K. K. & Levine, A. J. Identification of a novel p53 functional domain that is necessary for efficient growth suppression. *Proc. Natl. Acad. Sci. USA* **93**, 15335–15340 (1996).
40. Sakamuro, D., Sabbatini, P., White, E. & Prendergast, G. C. The polyproline region of p53 is required to activate apoptosis but not growth arrest. *Oncogene* **15**, 887–898 (1997).
41. Yu, H. *et al.* Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76**, 933–945 (1994).
42. Sun, X.-F. *et al.* A novel p53 germline alteration identified in a late onset breast cancer kindred. *Oncogene* **13**, 407–411 (1996).
43. Chittenden, T. *et al.* A conserved domain in bak, distinct from BH1 and BH2, mediates cell death and protein binding functions. *EMBO J.* **14**, 5589–5596 (1995).
44. Banks, L., Matlashewski, G. & Crawford, L. Isolation of human p53 specific monoclonal antibodies and their use in the study of human p53 expression. *Eur. J. Biochem.* **159**, 529–534 (1986).
45. Manos, M. M. *et al.* The use of polymerase chain reaction amplification for the detection of genital human papillomaviruses. *Cancer Cells* **7**, 209–214 (1989).
46. Snijders, P. J. *et al.* The use of general primers in the polymerase chain reaction permits the detection of a broad spectrum of human papillomavirus genotypes. *J. Gen. Virol.* **71**, 173–181 (1990).
47. Shamanin, V., Delius, H. & de Villiers, E.-M. Development of a broad spectrum PCR assay for papillomaviruses and its application in screening lung cancer biopsies. *J. Gen. Virol.* **75**, 1149–1156 (1994).
48. Chan, S.-Y., Delius, H., Halpern, A. L. & Bernard, H.-U. Analysis of genomic sequences of 95 papillomavirus types: uniting typing phylogeny and taxonomy. *J. Virol.* **69**, 3074–3083 (1995).

Acknowledgements. We thank P. Spink for assistance with the development of HPV screening and typing; I. Jacobs and H. Stevens for cervical and normal DNA; D. Pim for comments on the manuscript; and A. Norman for statistical analysis. G.M. acknowledges the support from the National Cancer Institute of Canada with funds from the Canadian Cancer Society and NSERC. C.H. is the recipient of a MRC Clinical Training Fellowship. G.M., A.S. and L.B. thank L. Crawford for insightful discussions.

Correspondence and requests for materials should be addressed to L.B. or G.M.

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

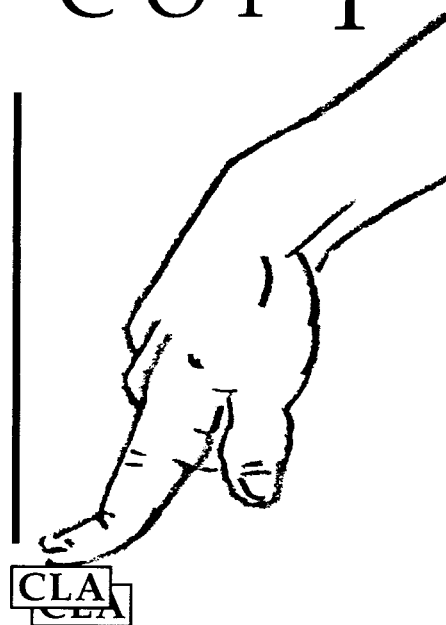
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised infringing photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986

An X-ray pulsar with a superstrong magnetic field in the soft γ -ray repeater SGR1806 – 20

C. Kouveliotou^{*†}, S. Dieters^{*‡}, T. Strohmayer^{†§}, J. van Paradijs^{‡||}, G. J. Fishman^{*}, C. A. Meegan^{*}, K. Hurley[¶], J. Kommers[#], I. Smith[☆], D. Frail^{**} & T. Murakami^{††}

^{*} ES-84, Space Science Laboratory, NASA/Marshall Space Flight Center, Huntsville, Alabama 35812, USA

[†] Universities Space Research Association

[‡] University of Alabama in Huntsville, Alabama 35899, USA

[§] Laboratory for High Energy Astrophysics, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

^{||} Astronomical Institute “Anton Pannekoek” & Center for High-Energy Astrophysics, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

[¶] Space Sciences Laboratory, University of California, Berkeley, California 94720-7450, USA

[#] Massachusetts Institute of Technology, MS 37-627, Cambridge, Massachusetts 02139, USA

[☆] Rice University, Space Physics and Astronomy, 6100 South Main, MS 108, Houston, Texas 77251-1892, USA

^{**} NRAO, VLA, 1003 Lopezville Road, Socorro, New Mexico 87801, USA

^{††} ISAS, 3-1-1 Yoshinodai, Sagami-hara, Kanagawa, Japan 229

Soft γ -ray repeaters (SGRs) emit multiple, brief (~ 0.1 -s), intense outbursts of low-energy γ -rays. They are extremely rare¹—three^{2–4} are known in our Galaxy and one⁵ in the Large Magellanic Cloud. Two SGRs are associated^{5–7} with young supernova remnants (SNRs), and therefore most probably with neutron stars, but it remains a puzzle why SGRs are so different from ‘normal’ radio pulsars. Here we report the discovery of pulsations in the persistent X-ray flux of SGR1806 – 20, with a period of 7.47 s and a spindown rate of $2.6 \times 10^{-3} \text{ s yr}^{-1}$. We argue that the spindown is due to magnetic dipole emission and find that the pulsar age and (dipolar) magnetic field strength are $\sim 1,500$ years and 8×10^{14} gauss, respectively. Our observations demonstrate the existence of ‘magnetars’, neutron stars with magnetic fields about 100 times stronger than those of radio pulsars, and support earlier suggestions^{8,9} that SGR bursts are caused by neutron-star ‘crust-quakes’ produced by magnetic stresses. The ‘magnetar’ birth rate is about one per millennium—a substantial fraction of that of

radio pulsars. Thus our results may explain why some SNRs have no radio pulsars.

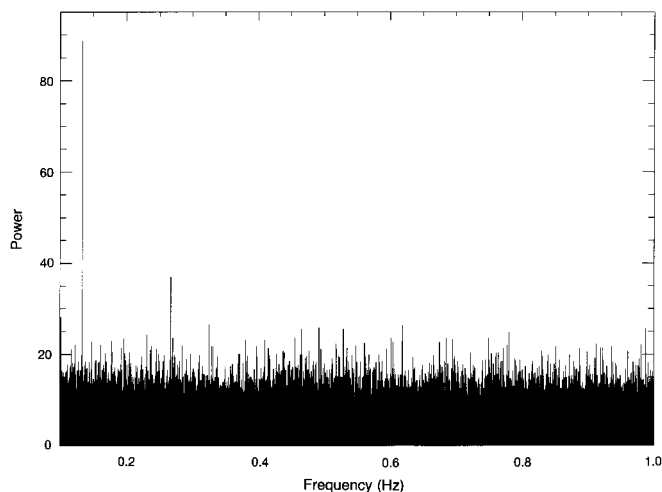
SGR1806 – 20 became extremely active between October 1996 and November 1997, when over 40 intense bursts and numerous weaker ones were detected¹⁰ with the Burst And Transient Source Experiment (BATSE) on board the Compton Gamma-Ray Observatory (CGRO). We observed SGR1806 – 20 with the Rossi X-Ray Timing Explorer (RXTE) five times between 5 and 18 November 1996, starting five days after the first triggered burst detection with BATSE. (Information on the archival data from RXTE/PCA and ASCA is available at <http://heasarc.gsfc.nasa.gov>.) During these observations¹¹, the source emitted series of outbursts in a ‘bunching’ mode, never seen before. The intensity of the outbursts, as well as the ‘bunching’ mode, varied significantly: mini-outbursts were interlaced with very intense ones and the rate of bursts varied from bunch to bunch (S. Dieters *et al.*, manuscript in preparation).

We made a period search of the data after excluding all bursts from the time series. The data were then energy-selected for 2–24 keV X-rays, background subtracted and binned at 0.5-s resolution. The resulting light curve was searched for periodicities between 0.03 and 1 Hz, by calculating a fast-Fourier-transform power spectrum (Fig. 1). The peaks in the spectrum are centred on the fundamental frequency of 0.13375 Hz (period of 7.47655 s) and its first harmonic at 0.26750 Hz. We find no significant power in any other frequency in the searched range. The probability that we detect a signal at the fundamental frequency this strong by chance coincidence is 1×10^{-13} (taking into account the number of trials, 1.9×10^6 , and the probability per trial, 5×10^{-20}).

To determine the fundamental period, all data sets were then corrected to the Solar System barycentre and separately folded at the longest detected period of 7.47655 s, and sub-harmonics thereof. These sub-harmonic folds showed multiple 7.47655-s pulses, which were identical, within statistical errors. We therefore conclude that 7.47655 s is the fundamental period.

To refine our period estimate, we have combined all Proportional Counter Array (PCA) data and used a ‘bootstrapping’ method, where the best period obtained from the shortest data span is used to count cycles and estimate the period over the next longest span of data, and so forth until a consistent period is found over the entire data set. This is achieved by cross-correlating the phase folds from data subsets using our best period estimate and fitting the resulting cross-correlation peaks with a gaussian. The phase offset is then converted to a time difference, a new period is calculated and the procedure is continued for the next (longest) span of data. Our errors in the period and period derivative were estimated by performing Monte Carlo simulations of 25 data sets, whereby we added gaussian noise with the data variance to the bins of each phase fold.

Figure 1 Fast-Fourier-transform power spectrum computed for the combined RXTE/PCA observations of 5/6, 7/13 and 17 November 1996 in the 2–24 keV range. (The data on 18 November contained too many bursts to be useful). The individual powers have been normalized by the average noise level to bring the distribution of noise powers into approximate agreement with the χ^2 distribution with 2 degrees of freedom. The data were sampled in 0.5-s bins. The 0.13375-Hz pulsed signal and its first harmonic at 0.26750 Hz stand out well above the noise level. The data have been barycentre corrected (using the co-ordinates of the LBV star mentioned in the text for the source position).



We find that a constant period does not fit the pulse arrival times well, and we derived an ephemeris of the pulse arrival times from the minimum in the phase-folded pulse light curve (see Fig. 2):

$$T_{\min}(\text{TJD}) = T_0 + NP + \frac{1}{2}P\dot{P}N^2$$

where T_0 is the reference time in truncated Julian days (TJD = JD - 2440000.5), P is the pulsar period, $\dot{P}$ is the period derivative and N is the number of elapsed cycles since T_0 . We find that $T_0 = 10392.450071037(1)$ (TJD), $P = 7.476551(3)$ s and $\dot{P} = 2.8(1.4) \times 10^{-11} \text{ s s}^{-1}$.

Using this ephemeris, the corresponding pulse profile of all the RXTE/PCA data is shown in Fig. 2a. The full amplitude of the pulsed flux is $1.0 \pm 0.2 \text{ counts s}^{-1}$ (between 2 and 24 keV). To estimate the point source count rate, we have folded the spectrum derived¹² from the observations of the Japanese satellite ASCA of the persistent X-ray source identified⁶ with SGR1806 - 20 through the detector response of the RXTE/PCA. It predicts $3.9 \text{ counts s}^{-1}$ for a point-like source detected with PCA in the 2-24 keV range; the corresponding pulsed flux amplitude is $\sim 26\%$.

We have searched the archival ASCA data of SGR1806 - 20 for the presence of a periodic signal over a period range that encompassed the original period and possible extrapolations given by our $\dot{P}$ measurement. A significant signal was found in the 1993 data set (epoch TJD 9275.85) at 7.46851 ± 0.00025 s with full amplitude of the pulsed flux $\sim 23\%$. The probability of this peak to appear by chance in the ASCA data is 3.6×10^{-4} , (taking into account the number of trials, 377, and the probability per trial, 9.66×10^{-7}). Figure 2b shows the pulse profile of the ASCA folded light curve. We have a marginal detection of the pulsation in the 1995 data set (7.4738 ± 0.001 s, epoch TJD 10006.72). The ASCA and RXTE/PCA periods are consistent with a long-term average spindown rate, $\dot{P} = (8.3 \pm 0.3) \times 10^{-11} \text{ s s}^{-1}$. The ASCA detection of the periodicity clearly associates the period with SGR1806 - 20 rather than with a hitherto unknown pulsar within the 1° field of view of the RXTE/PCA.

The period we have found in SGR1806 - 20 is very similar to the 8.0-s period found¹³ in SGR0526 - 66 during a three-minute interval following its extremely strong 5 March 1979 burst. This strongly suggests that the mechanism that produces SGR events is associated with a particular type of slowly rotating neutron stars.

What is the energy source of the SGR quiescence and burst emission? The (unabsorbed) flux (2-10 keV) of the quiescent emission from SGR1806 - 20 measured¹² with ASCA is $\sim 1 \times 10^{-11} \text{ erg cm}^{-2} \text{ s}^{-1}$. Assuming isotropic emission and a distance¹⁴ for SGR1806 - 20 of ~ 14 kpc, the 2-10 keV luminosity of the source is $L_{\text{SGR}} \approx 2 \times 10^{35} \text{ erg s}^{-1}$. Similarly, the SGR1806 - 20 burst peak luminosity¹⁵ (25-60 keV) corresponds to $\sim 10^{41} \text{ erg s}^{-1}$, which is $\geq 10^3$ times higher than the Eddington luminosity of a neutron star of 1.4 solar masses ($1.4 M_\odot$). Two types of energy sources have been proposed to explain the emission from SGRs: accretion, and energy released by the decay of the magnetic field of a strongly ($\geq 10^{14}$ G) magnetized neutron star, a so-called⁸ 'magnetar'. (Rotational energy loss, $\dot{E}_{\text{rot}}$, is not sufficient to explain the observed X-ray luminosity of SGR1806 - 20. For the observed P and $\dot{P}$ we find $\dot{E}_{\text{rot}} \approx 10^{33} \text{ erg s}^{-1}$ two orders of magnitude below L_{SGR}). As we argue below, the 'magnetar' model is the only one that can account for the observed properties of SGR1806 - 20.

SGR1806 - 20 has been identified with the SNR G 10.0 - 0.3, whose radio morphology is that of a plerion¹⁶ (a synchrotron nebula powered by the relativistic wind of a young pulsar); this suggests that it is centrally powered by a compact source. The X-ray spectrum¹² of the ASCA persistent source is a power law with photon index 2.2, and hydrogen column density, $N_{\text{H}} = 6 \times 10^{22} \text{ cm}^{-2}$. It does not show the emission lines commonly seen in SNRs, which shows that the contribution from shocked gas is relatively small and further supports the idea that it is an (X-ray) plerion. In addition, radio observations¹⁷ of SGR1806 - 20 show varying spatial structure at an angular scale of several arcseconds, perhaps a jet. All the above observations indicate that the SNR is powered by a wind of relativistic particles. Thompson and Duncan⁹ estimate that the particle luminosity, L_p , from SGR1806 - 20 is of the order of $10^{37} \text{ erg s}^{-1}$.

The coincidence of SGR1806 - 20 with an LBV (ref. 18), a rare type of evolved massive star, raises the possibility that the SGR persistent X-ray luminosity is caused by accretion from the LBV stellar wind. Accretion-powered emission, however, can be excluded by the following argument⁹, for reasonable (L. Kaper, personal communication) parameters of the LBV (wind velocity $V_w \approx 500 \text{ km s}^{-1}$, mass loss rate $\sim 10^{-5} M_\odot \text{ yr}^{-1}$, radius ~ 100 solar radii, and the distance a between the two stars $\geq 10^{13} \text{ cm}$): at the accretion radius ($r_{\text{acc}} = 2GM_{\text{NS}}/V_w^2$, where M_{NS} is the neutron

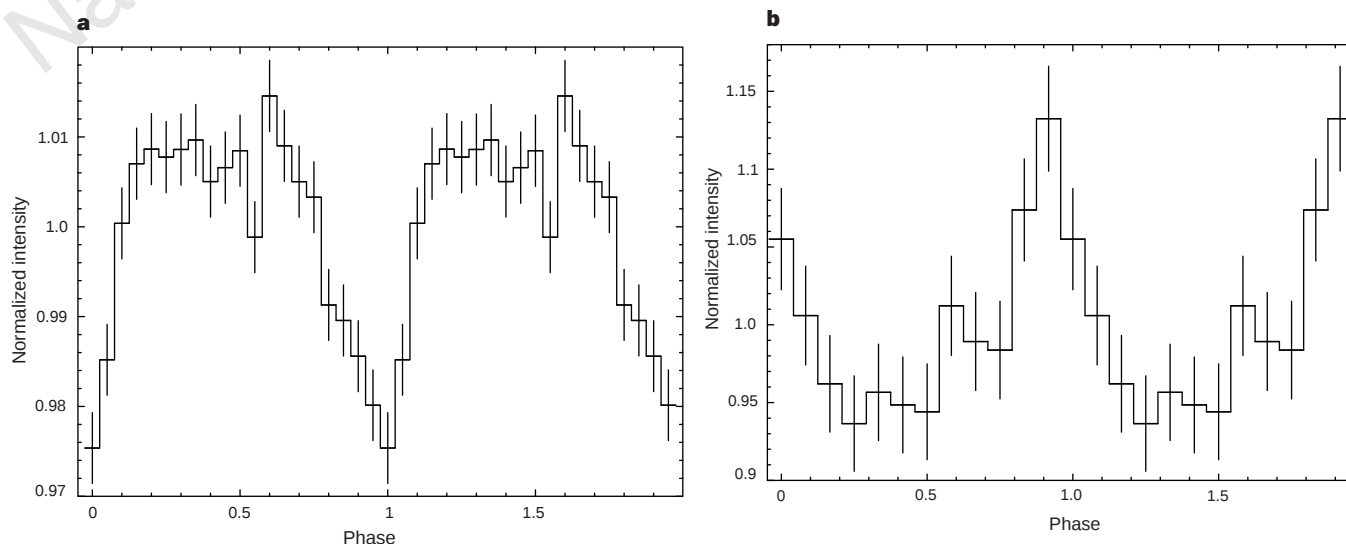


Figure 2 Epoch folded pulse profiles of SGR1806 - 20. **a**, RXTE/PCA data in 20 phase bins at the fundamental period of 7.476551 s, between 2 and 24 keV. **b**, ASCA data (from the 1993 observation of SGR1806 - 20) in 12 phase bins (to increase the signal to noise ratio) at the fundamental period of 7.4685125 s

between 2 and 10 keV. The plots in the two panels cover two periods, for clarity. The ASCA pulse profile is not derived by extrapolation of the RXTE/PCA period and period derivative; it is found by independent period searches.

star mass), the pressure exerted by the relativistic wind powering the plerion $P_{\text{wind}} = L_p / (4\pi r_{\text{acc}}^2)$ is $1.2 \times 10^3 \text{ dyn cm}^{-2}$. This pressure exceeds the stellar wind ram pressure ($P_{\text{ram}} = \rho_w V_w^2 \approx 25 \text{ dyn cm}^{-2}$, where ρ_w is the density in the wind) by almost two orders of magnitude. Thus, even if the SGR source were in a binary system with the LBV, accretion is highly unlikely to occur.

The radio data, however, may provide evidence for the orbital period of the system, that could be confirmed with a long-term monitoring of the source with RXTE. Frail *et al.*¹⁷ report that the jet-like structure associated with the plerion has rotated by 50° over an interval of 1.4 yr. It is possible that this structure is the result of a magnetotail of SGR1806 – 20 produced by the strong wind of the LBV. In such a case, we would expect the system orbital period to be of the order of 10 yr.

This leaves us with the magnetar model. Because in this model rotational losses do operate on the neutron star, we can estimate the spindown pulsar age, τ , and magnetic field, B , from its period and period derivative, according to¹⁹: $\tau = P/2\dot{P}$ and $B = 3.2 \times 10^{19} \sqrt{P\dot{P}}$, respectively. Substituting the above values for P and for the long-term $\dot{P}$ we find $\tau \approx 1,500 \text{ yr}$ and $B \approx 8 \times 10^{14} \text{ G}$. Thompson and Blaes²⁰ have shown that these formulae are of approximate validity only when the particle luminosity from a neutron star is much lower than its standard magnetic dipole luminosity. In particular, when the particle pressure is high enough to exceed the dipole pressure inside the light cylinder of the neutron star, the magnetic field lines will be combed out beyond a radius R_p , where $R_p/R_{\text{NS}} \approx (B_D^2 R_{\text{NS}}^2 c / 2L_p)^{1/4}$. Here B_D and R_{NS} are the dipole magnetic field and the neutron star radius, respectively. This raises the magnetic field strength at the light cylinder, and therefore the magnetic dipole torque. Following their analysis, we find for the magnetic field and spindown age of SGR1806 – 20, $\sim 2 \times 10^{14} \text{ G}$ and $\sim 8,000 \text{ yr}$, respectively. The latter value is in good agreement with an earlier age estimate¹⁶ of $\sim 10,000 \text{ yr}$ for G10.0 – 0.3 based on angular diameter versus surface brightness argument.

The magnetic field value is the highest measured so far (its uncorrected value is similar to that estimated²¹ for GB790305 and for²² 1E 1841 – 045 in the young SNR Kes 73); it supports the model for SGRs developed by Thompson and Duncan⁸. According to these workers, the SGR bursts are triggered by cracking of the neutron-star crust caused by magnetic stresses, which leads to sudden injection of Alfvén waves into the magnetosphere, particle acceleration, and formation of an optically thick pair plasma. Very high magnetic fields would explain the super-Eddington luminosities of SGRs, by suppressing the electron scattering cross-section of photons by a factor $\propto B^{-2}$. In addition, the decay of the magnetic field heats the neutron-star interior, giving rise to persistent thermal soft X-ray emission from the surface. Superstrong fields also lead to rapid spindown of a neutron star, which after a time interval of the order of 10^4 yr will rotate at a period in the 10-s range. Thus, unless G10.0 – 0.3 is not a plerionic SNR, the conclusion that SGR1806 – 20 is a magnetar seems unavoidable.

SGRs have several striking similarities with a small group of X-ray pulsars (only ~ 6 sources are known so far), called ‘anomalous’ X-ray pulsars (AXPs)^{23–25}, which are very different from normal binary X-ray pulsars, with respect to their spin period distribution (all between 6 and 12 s), their very soft X-ray spectra, and the absence of evidence for a binary companion. Several AXPs are associated with young SNRs; their very small distances to the galactic plane shows that AXPs are likewise recently formed neutron stars²⁴. Non-detection of orbital Doppler shifts in their X-ray pulsations indicates that if they have companions, these must be very low-mass stars.

The accretion model of AXPs fails to explain the narrow distribution of their observed pulse periods and X-ray luminosities. In the magnetar model the observed limited period range reflects

that the age range of AXPs is limited, and that for magnetic-dipole spindown the spin period varies rather slowly with time t as $t^{1/2}$. For a particular model of the B -field decay, Thompson and Duncan⁹ find for the period $P \approx (9 \text{ s}) B_{15} R_6^2 M_{1.4}^{-1/2} (t/10^4 \text{ yr})^{-1/2}$ and the X-ray luminosity $L_X \approx 5 \times 10^{34} (t/10^4 \text{ yr})^{-1/3} \text{ erg s}^{-1}$ (the latter expression depends somewhat on details of the neutron star equation of state). Here $B_{15} = B/10^{15} \text{ G}$, $R_6 = R_{\text{NS}}/10^6 \text{ cm}$, and $M_{1.4} = M_{\text{NS}}/1.4 M_\odot$. The observed X-ray luminosities and spin periods therefore follow naturally from the model.

According to a previous¹ analysis, the total number of SGRs in our galaxy is ≈ 7 . If the age estimate of SGR1806 – 20 is typical for these objects, their birth rate is of the order of one per millennium, that is, $\sim 10\%$ of the birth rate of normal radio pulsars^{26,27}. Neutron stars with superstrong magnetic fields may, therefore, constitute a non-negligible fraction of newly formed neutron stars and account for at least part of the discrepancy between the birth rates of core-collapse supernovae and of radio pulsars²⁸. It is unclear whether these magnetars represent the tail of a very broad neutron-star magnetic-field distribution, or whether they form a separate class of young neutron stars with superstrong magnetic fields.

We point out that the value of $P/\dot{P}$ of SGR1806 – 20 is smaller than any of the values (only five measured so far) for AXPs. This may suggest that SGRs are the earliest phase in the evolution of magnetars, followed by a phase in which they appear as AXPs. $\square$

Received 4 March; accepted 27 April 1998.

- Kouveliotou, C. *et al.* The rarity of soft γ -ray repeaters deduced from reactivation of SGR1806 – 20. *Nature* **368**, 125–127 (1994).
- Norris, J. P., Hertz, P., Wood, K. S. & Kouveliotou, C. On the nature of Soft Gamma Repeaters. *Astrophys. J.* **366**, 240–252 (1991).
- Hurley, K. in *Gamma-Ray Bursts: 3rd Huntsville Symposium* (eds Kouveliotou, C., Briggs, M. S. & Fishman, G. J.) 889–896 (AIP Conf. Proc. 384, Am. Inst. Phys., New York, 1995).
- Hurley, K. *IAU Circ. No.* 6743 (1997).
- Cline, T. L. *et al.* Precise source location of the anomalous 1979 March 5 gamma-ray transient. *Astrophys. J.* **255**, L45–L48 (1982).
- Murakami, T. *et al.* X-ray identification of the soft γ -ray repeater 1806-20. *Nature* **368**, 127–129 (1994).
- Kulkarni, S. R. & Frail, D. A. Identification of a supernova remnant coincident with the soft γ -ray repeater SGR1806 – 20. *Nature* **365**, 33–35 (1993).
- Thompson, C. & Duncan, R. The soft gamma repeaters as very strongly magnetized neutron stars. I. Radiative mechanism for outbursts. *Mon. Not. R. Astron. Soc.* **275**, 255–300 (1995).
- Thompson, C. & Duncan, R. The soft gamma repeaters as very strongly magnetized neutron stars. II. Quiescent neutrino, X-ray and Alfvén wave emission. *Astrophys. J.* **473**, 322–342 (1996).
- Kouveliotou, C. *et al.* *IAU Circ. No.* 6501 (1996).
- Kouveliotou, C. *et al.* *IAU Circ. No.* 6503 (1996).
- Sonobe, T., Murakami, T., Kulkarni, S. R., Aoki, T. & Yoshida, A. Characteristics of the persistent emission of SGR1806 – 20. *Astrophys. J.* **436**, L23–L25 (1994).
- Mazets, E. P., Golenetskii, S. V., Il'inski, V. N., Aptekar, R. L. & Guryan, Yu. A. Observations of a flaring X-ray pulsar in Dorado. *Nature* **282**, 587–589 (1979).
- Corbel, S. *et al.* The distance of the soft gamma repeater SGR1806 – 20. *Astrophys. J.* **478**, 624–630 (1997).
- Kouveliotou, C. SGR activity and global characteristics. *Astrophys. Space Sci.* **231**, 49–56 (1995).
- Kulkarni, S. R., Frail, D. A., Kassim, N. E., Murakami, T. & Vasisht, G. The radio nebula of the soft γ -ray repeater 1806 – 20. *Nature* **368**, 129–131 (1994).
- Frail, D. A., Vasisht, G. & Kulkarni, S. R. The changing structure of the radio nebula around the soft gamma-ray repeater SGR1806 – 20. *Astrophys. J.* **480**, L129–L132 (1997).
- Van Kerkwijk, M. H., Kulkarni, S. R., Matthews, K. & Neugebauer, G. A luminous companion to SGR1806 – 20. *Astrophys. J.* **444**, L33–L35 (1995).
- Michel, F. C. *Theory of Neutron Star Magnetospheres* (Univ. Chicago Press, 1991).
- Thompson, C. & Blaes, O. Magnetohydrodynamics in the extreme relativistic limit. *Phys. Rev. D* **57**(6), 3219–3234 (1998).
- Paczynski, B. GB790305 as a very strongly magnetized neutron star. *Acta Astron.* **42**(3), 145–153 (1992).
- Vasisht, G. & Gotthelf, E. V. The discovery of an anomalous X-ray pulsar in the supernova remnant Kes 73. *Astrophys. J.* **486**, L129–L132 (1997).
- Mereghetti, S. & Stella, L. The very low mass X-ray binary pulsars: a new class of sources? *Astrophys. J.* **442**, L17–L20 (1995).
- Van Paradijs, J., Taam, R. E. & van der Heuvel, E. P. J. On the nature of the ‘anomalous’ 6-s X-ray pulsars. *Astron. Astrophys.* **299**, L41–L44 (1995).
- Duncan, R. C. & Thompson, C. in *High Velocity Neutron Stars and Gamma-Ray Bursts* (eds Rothschild, R. E. & Lingenfelter, R. E.) 111–117 (AIP Conf. Proc. 366, Am. Inst. Phys., New York, 1995).
- Lyne, A. G. & Graham-Smith, F. *Pulsar Astronomy* (Cambridge Univ. Press, 1990).
- Narayan, R. & Ostriker, J. P. Pulsar populations and their evolution. *Astrophys. J.* **352**, 222–246 (1990).
- Frail, D. A. in *The Many Faces of Neutron Stars* (eds Buccieri, R., van Paradijs, J. & Alpar, M. A.) (Kluwer Academic, in the press).

Acknowledgements. We thank R. Duncan, C. Thompson and M. Finger for discussions.

Correspondence and requests for materials should be addressed to C.K. (e-mail: chryssa.kouveliotou@msfc.nasa.gov).

Irregular variations in the melting point of size-selected atomic clusters

Martin Schmidt, Robert Kusche, Bernd von Issendorff & Hellmut Haberland

Fakultät für Physik, Universität Freiburg, H. Herderstrasse 3, D-79104 Freiburg, Germany

Small particles have a lower melting point than bulk material¹. The physical cause lies in the fact that small particles have a higher proportion of surface atoms than larger particles—surface atoms have fewer nearest neighbours and are thus more weakly bound and less constrained in their thermal motion^{2,3} than atoms in the body of a material. The reduction in the melting point has been studied extensively for small particles or clusters on supporting surfaces. One typically observes a linear reduction of the melting point as a function of the inverse cluster radius^{2,4,5}. Recently, the melting point of a very small cluster, containing exactly 139 atoms, has been measured in a vacuum using a technique in which the cluster acts as its own nanometre-scale calorimeter^{6,7}. Here we use the same technique to study ionized sodium clusters containing 70 to 200 atoms. The melting points of these clusters are on average 33% (120 K) lower than the bulk material; furthermore, we observe surprisingly large variations in the melting point (of ± 30 K) with changing cluster size, rather than any gradual trend. These variations cannot yet be fully explained theoretically.

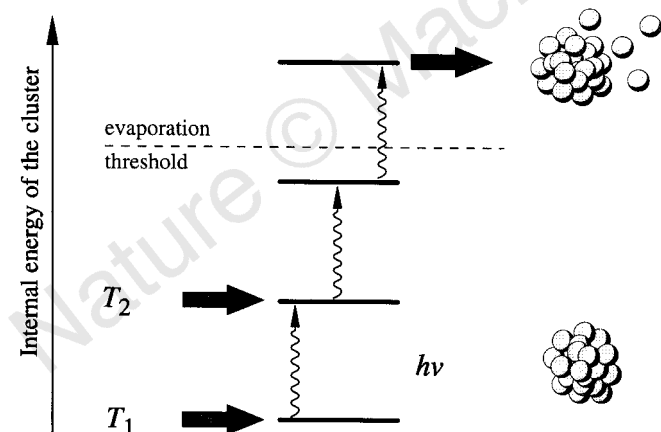


Figure 1 The basic idea of the experiment is to heat a cluster at temperature T_1 by absorbing a photon of energy $\delta U = h\nu$. The increased temperature of the cluster, T_2 , is then identified by increasing the cluster source temperature until the thermally heated clusters show the same photofragmentation behaviour as the laser-heated ones. This procedure can be divided into four steps. (1) A mass-selected beam of cluster ions of temperature T_1 is prepared in vacuum. (2) The clusters are irradiated by photons from a laser. The photon energy ($h\nu = 3.1 \text{ eV} \approx 10^{-18} \text{ J}$) quickly relaxes into vibrations and heats the cluster to a temperature T_2 , where the clusters do not emit atoms on the 100- μs timescale of the experiment. Only the absorption of more photons from the same laser pulse (two are shown here) raises the temperature above T_{evap} , the temperature needed for evaporation of atoms from the cluster. (3) The size distribution of the remaining cluster ions is measured, which is a very sensitive measure of the cluster's internal energy. (4) The cluster temperature is now raised to T_2 where the absorption of only two photons leads to the same number of evaporated atoms. An energy increase δU thus leads to a temperature increase $\delta T = T_2 - T_1$. For small enough δT , the ratio $\delta U/\delta T$ equals the heat capacity $c(T)$. In our experiment this is the case, as discussed in more detail in ref. 6.

The melting point (T_{melt}) of small particles is not only of scientific interest, but also has some technological implications. In sintering processes, fine powders are compressed and heated until they coalesce. If extremely fine powders are employed, a lower sintering temperature could be used. Also, the present drive towards nanoscale technology leads to smaller and smaller geometric dimensions with a concomitant reduction of T_{melt} and consequently reduced electrical and mechanical stability at elevated temperatures.

The standard way to measure T_{melt} of a material is to heat it and record the temperature at which it becomes liquid. In order to do this, one needs some physical property which changes measurably at T_{melt} . For example, the electron-diffraction pattern of a crystalline solid, which vanishes on melting, has been used to study the melting of small particles supported on surfaces^{4,5} or in cluster beams⁸. In both cases one has to work with broad size distributions. Moreover, in the case of the surface experiments the physical properties of the clusters could be affected by the surface contact. There exist two earlier experiments on the melting of free, size-selected clusters: one studied the temperature dependence of the ionization energy of sodium clusters⁹, the other looked for a transition of methanol hexamers¹⁰.

A deeper insight into the solid-to-liquid phase transition may be gained by measuring the relation between temperature and internal energy across the melting point. For a macroscopic material this is achieved by placing the sample in a thermally insulated box, containing an electric heater and a thermometer. Some known amount of energy U is supplied to the material by heating, and its temperature T is measured. The relation between temperature and energy, $U = U(T)$, is called the caloric curve, its derivative is the heat capacity $c(T) = \partial U/\partial T$, and the equipment used to measure it is called a calorimeter. At the melting point, there is a step in the caloric curve (as shown in Fig. 2), which is simple to understand. We

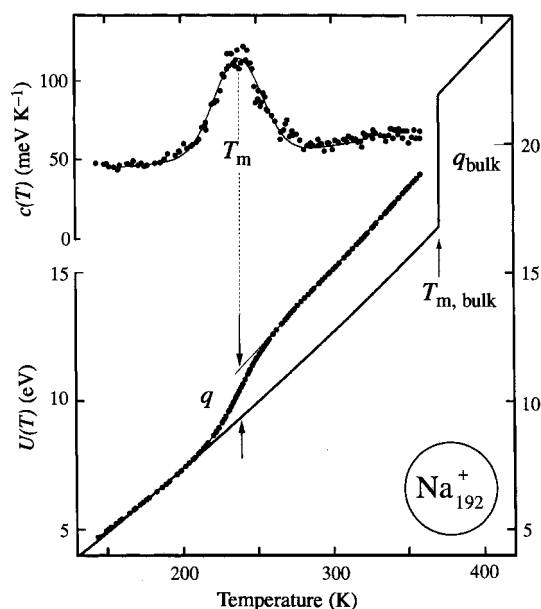


Figure 2 Heat capacity $c(T)$ and its integral, the caloric curve $U(T)$, for a positively charged sodium cluster containing exactly 192 atoms. The solid dots show the experimental results. From the maximum of $c(T)$ one obtains the melting point T_{melt} , and from the height of the smoothed-out step in the caloric curve one gets the latent heat of fusion q , that is, the energy necessary to destroy the crystalline lattice at the melting point. The thick solid line is the bulk caloric curve scaled to 192 atoms. $T_{\text{m,bulk}}$ and q_{bulk} are the bulk melting point and latent heat of fusion, respectively. Heat-capacity curves have been measured for sodium clusters with 70–200 atoms. Melting points and latent heats have been extracted and are shown in Fig. 3b and c.

consider a piece of ice floating in water, in thermal equilibrium at a temperature of 0 °C. Heating this system with a small amount of energy would not lead to a temperature increase: instead, the energy would be used to melt the ice. Hence the energy increases without a temperature increase, which corresponds to a vertical line in the caloric curve. The height of the step is equal to the energy needed to destroy the crystalline lattice at T_{melt} . This energy is called the latent heat of fusion q .

A large sample mass is required for such a calorimeter in order to minimize the effects of the thermal capacities of the heater and thermometer. The limit of the technologically possible is reached when the heater and thermometer are the same supporting micro-structure. This has been achieved recently in an experiment on supported tin clusters⁵.

The ultimate miniaturization of the calorimeter is reached if one dispenses entirely with supporting surface, heater and thermometer. It was shown earlier^{6,7} that this is indeed possible, by measuring the internal energy of free clusters with a variable temperature. Clusters can be thermalized by embedding them for some time in an inert buffer gas of known temperature^{6,9,11}. To measure their internal energy, known amounts of energy are added by irradiating the clusters with a laser. By counting the number of evaporated atoms one can determine the internal energy of the cluster before the irradiation. Such a procedure, however, still requires input parameters like the atom binding energies, which are not known accurately enough. We have instead used a differential measuring scheme, which does not require any parameters except for the photon energy and the temperature of the clusters before they absorb a photon. This is explained in Fig. 1 and in more detail in refs 6 and 7.

One example of a caloric curve that we obtained in this way is given in Fig. 2. The caloric curves of small particles do not have a sharp step at T_{melt} , but show a more gradual increase. The finite width of the step is due to the finite number of atoms in the cluster²⁻⁴; only in the limit of very many atoms does the increase become step-like. We measured caloric curves for sodium clusters containing between 70 and 200 atoms. The corresponding melting points and latent heats are given in Fig. 3b and c. Neither value shows the expected gradual increase to the bulk value, but rather shows a pronounced structure, which to our knowledge has not been experimentally observed before. Changing the cluster size by just one atom can shift the melting point by several degrees.

Similar variations of the melting point have been predicted in a theoretical treatment of argon clusters¹². Argon clusters are very simple systems, in which the atoms tend to organize in onion-like shells of icosahedral symmetry. Completion of shells occurs at cluster sizes $N = 13, 55, 147, 309 \dots$; for these "magic numbers" local maxima of T_{melt} were found, which can be explained by the enhanced rigidity of argon clusters with completed shells.

The question arises whether such a simple argument is appropriate also for the observed behaviour of our sodium cluster melting points. Depending on size, the stability of sodium clusters can be explained using two different concepts: electronic and geometric shell closings^{9,13,14}. Sodium clusters are surprisingly well described by the 'nearly free electron model', in which the valence electrons are treated as nearly free, being caged only by the cluster's surface^{13,14}. The electrons are grouped in shells (similar to what is known for atoms or nuclei), and on closing such a shell, the electronic contribution to the binding energy of an atom reaches a maximum^{13,14}. This is demonstrated by the mass spectrum of hot sodium clusters shown in Fig. 3a. The maxima (magic numbers) correspond to clusters with closed electronic shells. They have a larger binding energy, and are thus more resistant to evaporation.

Could an increased binding energy per atom lead to an increased melting point? The mass spectrum shows the stability with respect to evaporation, where an atom and its valence electron leave the cluster. The loss of one electron requires a lot of energy for a

closed-shell cluster, which therefore has a high binding energy. On melting, conversely, the number of valence electrons stays the same. One can thus expect that electronic shells should be more visible in the fragmentation pattern than in the melting process. Nevertheless, there are some similarities. There is a maximum near Na_{93}^+ in both cases. Also, near 139 atoms, where there is the next electronic shell closing, there is a maximum in T_{melt} . However, a more detailed examination of this size region showed that the peak of T_{melt} occurs at 142 atoms, where the intensity of the mass spectrum (and thus the binding energy of the clusters) has already decreased. The differences become more marked at the next electronic shell closing near 197 atoms, where there are no special features in the melting point. The effects of the shell closing on the size dependence of the latent heat q are even less marked. The electronic structure of the clusters

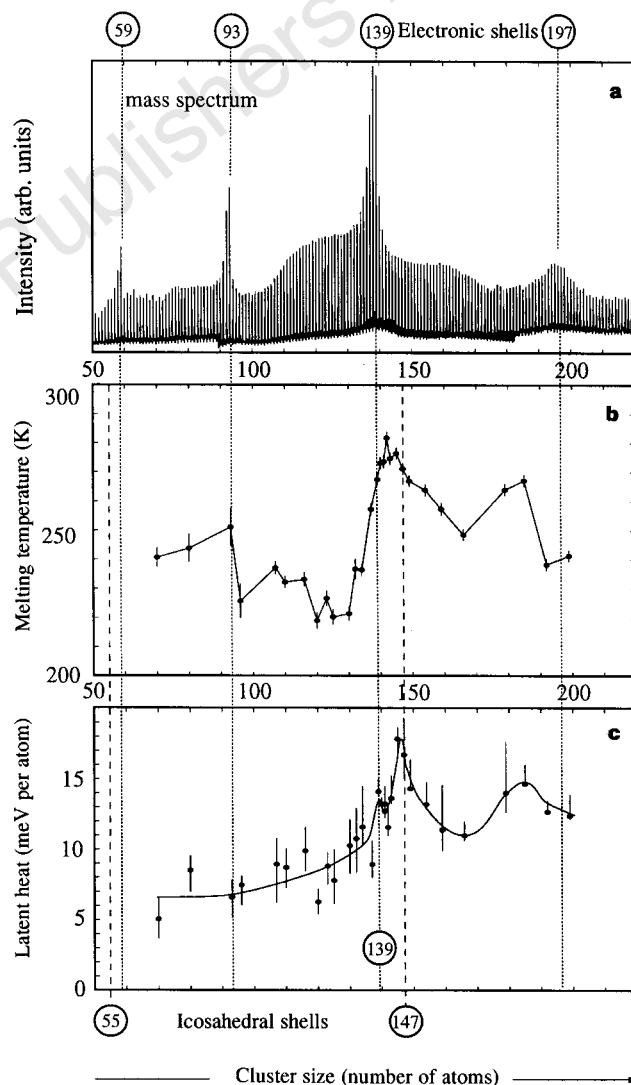


Figure 3 Three different data sets plotted against the cluster size. **a**, Mass spectrum of hot, positively charged sodium clusters ($T \approx 400$ K). The enhanced intensities of the so-called "magic" cluster sizes can be explained by higher dissociation energies at the electronic shell closings^{13,14} indicated above the figure. These numbers (59, 93, ...) correspond to the number of atoms in the cluster. One obtains the number of valence electrons by subtracting one unit. **b**, Melting-point temperatures. The values change strongly with cluster size and show some weak correlation with electronic shell closings. The sodium bulk value (371 K) is far above the cluster melting points. **c**, Latent heat of fusion. The bulk value is 27 meV per atom. The number of atoms where icosahedral shell closings⁹ occur are indicated below the figure.

can thus explain the magic numbers of the mass spectrum, but not the size-dependence of T_{melt} and q .

For large sodium clusters of several thousand atoms, geometrical shell closings can explain the intensities in a mass spectrum⁹. But smaller clusters could also have an icosahedral structure, and although this does not show up in the mass spectrum, it might well have an influence on the cluster melting points. There is an icosahedral shell closing at 147 atoms, and indeed the latent heat has a pronounced maximum in this region. However, it is not possible to explain all the observed features by geometrical models, either by the icosahedral structure mentioned above, or by the modified pentagonal bipyramid structures favoured by gold clusters in this size range¹⁵. Thus geometrical shell closing arguments cannot explain the size dependence of either T_{melt} or q .

We therefore conclude that the size dependence of melting-point temperature and latent heat of fusion cannot be explained by one simple argument. There is probably a complicated interplay between geometric and electronic structure, presenting a challenge for theory. □

Received 8 December 1997; accepted 16 February 1998.

1. Pawlow, P. Über die Abhängigkeit des Schmelzpunktes von der Oberflächenenergie eines festen Körpers. *Z. Phys. Chem.* **65**, 1–35 (1909).
2. Couchman, R. R. The Lindemann hypothesis and the size dependence of melting temperatures II. *Phil. Mag. A* **40**, 637–643 (1979).
3. Berry, R. S. When the melting and freezing point are not the same. *Sci. Am.* **263**, 50–56 August (1990).
4. Buffat, Ph. & Borel, J. P. Size effect on the melting temperature of gold clusters. *Phys. Rev. A* **13**, 2287–2298 (1976).
5. Lai, S. L., Guo, J. Y., Petrova, V., Ramanath, G. & Allen, L. H. Size-dependent melting properties of small tin particles: noncalorimetric measurements. *Phys. Rev. Lett.* **77**, 99–102 (1996).
6. Schmidt, M., Kusche, R., Kronmüller, W., v. Issendorff, B. & Haberland, H. Experimental determination of the melting point and heat capacity for a free cluster of 139 sodium atoms. *Phys. Rev. Lett.* **79**, 99–102 (1997).
7. Bertsch, G. Melting in clusters. *Science* **277**, 1619 (1997).
8. Hovick, J. W. & Bartell, L. S. Phase and phase-changes in tert-butyl thiol. *J. Mol. Struct.* **413**, 615–620 (1997).
9. Martin, T. P. Shells of atoms. *Phys. Rep.* **273**, 199–241 (1996).
10. Buck, U. & Eitsscher, I. Experimental evidence for an isomeric transition of size selected methanol hexamers. *J. Chem. Phys.* **100**, 6974–6976 (1994).
11. Ellert, Ch., Schmidt, M., Schmitt, Ch., Reiners, Th. & Haberland, H. Temperature dependence of the optical response of small, open shell sodium clusters. *Phys. Rev. Lett.* **75**, 1731–1734 (1995).
12. Wales, D. J. & Berry, R. S. Freezing, melting, spinodals, and clusters. *J. Chem. Phys.* **92**, 4473–4481 (1990).
13. de Heer, W. A. The physics of simple metal clusters: experimental aspects and simple models. *Rev. Mod. Phys.* **65**, 611–676 (1993).
14. Brack, M. The physics of simple metal clusters: self-consistent jellium model and semi-classical approaches. *Rev. Mod. Phys.* **65**, 677–732 (1993).
15. Cleveland, C. L. *et al.* Structural evolution of smaller gold nanocrystals: the truncated decahedral motif. *Phys. Rev. Lett.* **79**, 1873–1876 (1997).

Acknowledgements. This work was supported by the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to H.H. (e-mail: hellmut@frha06.physik.uni-freiburg.de).

Carbon nanotubes as long ballistic conductors

C. T. White*† & T. N. Todorov*

* Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, UK

Early theoretical work on single-walled carbon nanotubes^{1–3} predicted that a special achiral subset of these structures known as armchair nanotubes³ should be metallic. Tans *et al.*⁴ have recently confirmed these predictions experimentally and also showed directly that coherent electron transport can be maintained through these nanowires up to distances of at least 140 nm. But single-walled armchair nanotubes are one-dimensional conductors with only two open conduction channels (energy subbands in a laterally confined system that cross the Fermi level)^{1–3}. Hence, with increasing length, their conduction electrons ultimately become localized⁵ owing to residual disorder in the tube which is inevitably produced by interactions between the tube and its environment. We present here calculations which show, however, that unlike normal metallic wires, conduction electrons in armchair nanotubes experience an effective disorder averaged over the tube's circumference, leading to electron mean free paths that increase with nanotube diameter. This increase should result in exceptional ballistic transport properties and localization lengths of 10 μm or more for tubes with the diameters that are typically produced experimentally⁶.

Once physisorbed on a surface, even an initially perfect metallic carbon nanotube is disordered because of residual interactions with the substrate. In general, for such long thin conductors at zero temperature, the theory of transport through weakly disordered materials predicts a transition to a localized regime (where the resistance R is exponentially large) around a wire length where R reaches a value of one quantum resistance unit⁷, $r = h/2e^2$. Deep within the localized regime the resistance increases exponentially with length L as $R(L) \sim e^{L/\xi}$, where ξ is the localization length for electrons in the disordered wire⁸. For wires with N_C open conduction channels, ξ is given within a factor of order unity by $\xi = N_C l$ (refs 9, 10) where l is the elastic mean free path for backward electron scattering. Hence for wires with $N_C \gg 1$, the motion of the electron on the scale of ξ is largely diffusive, but for weakly disordered, small-diameter armchair nanotubes, which have $N_C = 2$, this motion is largely ballistic. This analysis suggests immediately that any observable quantum transport through these nanotubes is also ballistic.

For a fixed amount of disorder, as the transverse size of a normal metallic quantum wire increases l remains largely fixed but ξ increases due to the introduction of new channels at the Fermi level, ϵ_F (see ref. 10 and refs therein). However, this behaviour does not lead to long localization lengths in armchair carbon nanotubes because N_C remains pinned at two. We will show, however, that for a fixed amount of disorder as the radius of a small-diameter armchair tube increases, l does not remain fixed but rather increases leading to long localization lengths. We will also show that this remarkable behaviour arises because of the special character of the armchair states close to ϵ_F in the perfect tube and hence does not depend crucially on the details of the disorder.

To study the effects of disorder on the transport properties of armchair nanotubes we adopt the usual tight-binding model which retains only the nearest neighbour π -like hamiltonian matrix elements between $|p_\perp\rangle$ orbitals (one per carbon atom) orientated normal to the tubule surface^{1,3}. Local density-functional calculations have established that this model, with all diagonal matrix elements fixed at ϵ_F and all non-zero off-diagonal matrix elements fixed at $V_0 = -2.7\text{ eV}$ provides an excellent description of the valence bands of perfect armchair carbon nanotubes in the vicinity of ϵ_F (refs 1, 11). The effects of disorder can then be incorporated into the model (referred to here as the 'full model') by assuming that these diagonal and off-diagonal matrix elements are not fixed at their values in the perfect tube but are independent random variables with variances σ_e^2 and σ_v^2 , respectively.

In the absence of disorder, the full model yields $2N_B$ bands as a function of the quasi-momentum k that labels an eigenstate of the helical screw operator $S(\theta_0, z_0)$ used to generate the $[N_B, N_B]$ armchair tube by starting with a single ring, such as highlighted in Fig. 1a, and then rotating this ring by $\theta_0 = \pi/N_B$ radians around the tubule axis followed by a translation $z_0 = (\sqrt{3}/2)d_0$ along this axis¹¹. Each ring contains $2N_B$ carbon atoms and has a radius $r_T = (3d_0/2\pi)N_B$, where N_B is the number of C–C bonds in the ring and $d_0 = 0.142\text{ nm}$ is the C–C bond distance. The a_1 and a_2 bands that cross at $k_F = 2\pi/3$ are present in all armchair tubules¹. These two bands are highlighted in Fig. 1b for the [10,10] tube. In the absence of disorder, symmetry can be used to block diagonalize the full model to the point that the a_1 and a_2 bands are described

† Permanent address: Code 6179, Naval Research Laboratory, Washington DC 20375, USA.

exactly (within an unimportant constant) by the hamiltonian:

$$H_0 = V_0 \sum_{j=1}^2 (-1)^{(j+1)} \left[\sum_m [|a_{jm}\rangle \langle a_{jm}| + (|a_{jm}\rangle \langle a_{j(m+1)}| + \text{H.c.})] \right] \quad (1)$$

where $|a_{1m}\rangle$ ($|a_{2m}\rangle$) denotes the a_1 (a_2) state constructed from rotationally invariant—under C_{N_B} rotations—symmetric (anti-symmetric) sums of $|p_\perp\rangle$ orbitals for each of the N_B pairs of nearest neighbours in the $2N_B$ atom carbon ring labelled m and H.c. is the Hermitian conjugate.

Breaking the symmetries used to derive equation (1), disorder introduces additional couplings not only between the states making up the a_1 and a_2 bands but also between these states and states making up the other $2(N_B - 1)$ bands present in the full model. However, for small-diameter tubes these other bands are separated by at least $\pm \pi V_0/N_B$ from ϵ_F . Hence, if the disorder is not too severe and r_T not too large, we should still be able to work entirely within the reduced basis that describes the a_1 and a_2 bands. With this approximation we arrive at a two-band model for studying the transport properties of small-diameter, weakly disordered armchair nanotubes at and around ϵ_F . The hamiltonian describing this reduced model is given by $H = H_0 + U$, where U is the projection onto the reduced two-band basis of the perturbation introduced by the disorder into the full model.

In its most general form H is given by:

$$H = \sum_m \left\{ \sum_{j=1}^2 \left\{ (\bar{\epsilon}_m + (-1)^{(j+1)} \hat{V}_m) |a_{jm}\rangle \langle a_{jm}| + [(-1)^{(j+1)} \bar{V}_m |a_{jm}\rangle \langle a_{j(m+1)}| + \text{H.c.}] \right\} + [\bar{\epsilon} |a_{1m}\rangle \langle a_{2m}| + \bar{V}_m (|a_{1m}\rangle \langle a_{2(m+1)}| - |a_{2m}\rangle \langle a_{1(m+1)}|) + \text{H.c.}] \right\} \quad (2)$$

where $\bar{\epsilon}_m = (1/N_B) \sum_{n=1}^{N_B} (V_{mmnn}^{11} + V_{mmnn}^{22})/2$; $\bar{\epsilon}_m = (1/N_B) \sum_{n=1}^{N_B} (V_{mmnn}^{11} - V_{mmnn}^{22})/2$; $\hat{V}_m = (1/N_B) \sum_{n=1}^{N_B} V_{mmnn}^{12}$; $\bar{V}_m = (1/N_B) \sum_{n=1}^{N_B} (V_{mn(m+1)n}^{21} + V_{mn(m+1)(n-1)}^{12})/2$; $\bar{V}_m = (1/N_B) \sum_{n=1}^{N_B} (V_{mn(m+1)n}^{21} - V_{mn(m+1)(n-1)}^{12})/2$, and the prefactors of $1/N_B$ arise from the normalization of the ring states. The matrix elements V_{pqmn}^{ij} are defined by $V_{pqmn}^{ij} = \langle pq; i | H_T | mn; j \rangle$, where $|mn; 1\rangle$ ($|mn; 2\rangle$) denotes the $|p_\perp\rangle$ orbital associated with the carbon labelled 1 (2) in the C-C bond labelled n modulo N_B in the ring labelled m . In this notation the C-C bond defined by n and m' is located by rotating and translating the two carbon atoms in the C-C bond defined by n and m by $(m' - m)$ applications of $S(\theta_0, z_0)$. Finally, H_T denotes the full model hamiltonian in the presence of disorder. Hence the disorder introduced into H_T manifests itself in H as averages of the matrix elements of H_T over the number of carbon bonds in a ring.

For the two-band model in the weak scattering limit l is given with the aid of Fermi's golden rule by:

$$\frac{1}{l} = \frac{2\pi}{v_F \hbar} \left[\frac{1}{2} \sum_{j,j'=1}^2 \left\langle \left| k_{jF} [H - \langle H \rangle] (-1)^{(j-j'+1)} k_{j'F} \right|^2 \right\rangle \right] \rho(\epsilon_F), \quad (3)$$

where l is measured in number of rings along the tube, $v_F = \sqrt{3} |V_0| / \hbar$, $\langle \dots \rangle$ denotes an ensemble average over the disordered potentials, $|k_{jF}\rangle = (1/\sqrt{N}) \sum_m e^{imk_F} |a_{jm}\rangle$ with $k_F = 2\pi/3$, H is defined by equation (2), and $\rho(\epsilon_F) = N/2\pi \sqrt{3} |V_0|$ is the density of back scattered states at ϵ_F for either the a_1 or a_2 bands for a tube containing N rings in the limit of large N . Equation (3) yields:

$$l = \frac{6V_0^2}{(2\sigma_\epsilon^2 + 9\sigma_V^2)} N_B. \quad (4)$$

In arriving at equation (4) we have repeatedly used the relation $\langle (1/N_B^2) \sum_{i,j=1}^{N_B} [(x_i - x_0)(x_j - x_0)] \rangle = \sigma_x^2/N_B$ valid for independent random variables x_i with average x_0 and variance σ_x^2 . The derivation

of equation (4) requires that the two-band model is valid and hence that the disordered diagonal and off-diagonal matrix elements of H have standard deviations ($\sigma_\epsilon/\sqrt{N_B}$, $\sigma_V/\sqrt{N_B}$) that are small with respect to $\frac{\pi |V_0|/N_B}{3|V_0|d_0/2r_T}$ or equivalently $\max(\sigma_\epsilon, \sigma_V) \ll \sqrt{3\pi d_0/2r_T} |V_0|$. Hence, if the disorder—measured by $\max(\sigma_\epsilon, \sigma_V)$ —is not too severe and r_T not too large, equation (4) should be valid. But for small-diameter residually disordered tubes this constraint, which prevents the a_1 and a_2 states at ϵ_F from mixing appreciably with the other bands of the full model, should be satisfied. In particular, for the [10,10] tube it becomes, $\max(\sigma_\epsilon, \sigma_V) \ll \pi |V_0|/\sqrt{10} = 2.7 \text{ eV}$.

Equation (4) predicts that weakly disordered small-diameter armchair tubules differ fundamentally from normal metallic wires which have electron mean free paths along the wire independent of the wire's transverse size. In a normal metallic wire the disorder-induced variance of the hamiltonian matrix elements between transverse wire states will contain a factor analogous to N_B^{-1} for the tube. However, in such a wire the number of open conduction channels, N_C , at ϵ_F will grow as N_B and hence so too will the total density of states, making l independent of the wire's transverse size. In contrast, for armchair nanotubes, N_C is fixed at two, the density

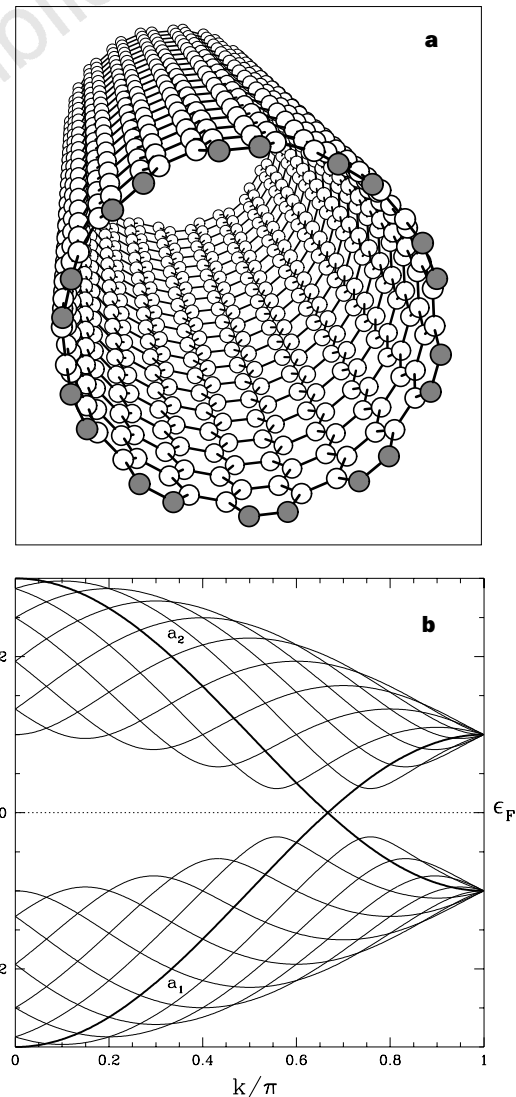


Figure 1 Geometry and band structure of the [10,10] armchair nanotube. **a**, A segment of the [10,10] armchair tubule with a planar 20-atom carbon ring containing 10 C-C bonds that can be used to generate the entire tubule depicted in grey. **b**, Band structure of the perfect [10,10] tubule within the full model. The a_1 and a_2 bands which cross at $\epsilon_F (=0.0 \text{ eV})$ are depicted as heavy lines.

of states does not contain the compensating factor for N_B^{-1} , and l will grow as N_B . Equation (4) also predicts the small-diameter, weakly disordered armchair nanotubes—with N_C pinned at 2—will have localization lengths (given by $\xi \approx N_C l$) that are comparable to those in similarly disordered normal wires with similar Fermi velocities but with $N_C = 2N_B$. Finally, equation (4) predicts that residually disordered tubes will show ballistic electron transport over unprecedented distances for such laterally confined systems. This follows because the chemical and mechanical stability of the tube make σ_ϵ^2 and σ_V^2 small, the strong C–C interaction makes $|V_0|$ large, and the special character of the armchair states at ϵ_F enhances l by a factor of N_B .

As an example of the last prediction, we consider a positionally disordered tube ($\sigma_\epsilon = 0$ eV, $\sigma_V \neq 0$ eV) where the C–C bond lengths vary according to a rectangular probability distribution by as much as $\pm 1\%$ of d_0 . Because the off-diagonal matrix elements of the full model should change with C–C bond length as in polyacetylene¹ ($\delta V = \alpha \delta d$ with $\alpha \approx 47$ eV nm^{−1}), σ_V will be ~ 0.04 eV. Equation (4) then yields $l = 3.7$ μ m ($\xi \approx 7.5$ μ m) for $V_0 = -2.7$ eV and $N_B = 10$. As an additional example, we consider a substitutionally disordered tube ($\sigma_\epsilon \neq 0$ eV, $\sigma_V = 0$ eV) where the probability of finding an impurity with potential ϵ_a is given by P . If the impurities correspond to nitrogen atoms ($\epsilon_a = -2.5$ eV; ref. 12) separated on average by 10 nm along the [10,10] tube ($P = 6.0 \times 10^{-4}$), then $\sigma_\epsilon = \sqrt{P(1-P)}|\epsilon_a| = 0.06$ eV and equation (4) yields $l = 7.5$ μ m ($\xi \approx 15$ μ m) for $V_0 = -2.7$ eV and $N_B = 10$.

Substrate interactions are not expected to significantly flatten tubes such as the [10,10] (ref. 13). In addition, twists along the tube are unlikely to be produced by its van der Waals interaction with the substrate. On the other hand, bends caused by nanometre- (or larger-) scale substrate features can occur and could limit l (ref. 4). The principal effect of a bend will be to shift the inter-ring matrix elements contributing to $\bar{V}_m$ and V_m in equation (2). However, for every such matrix element in the full model that is increased by a small amount by the bend there is a corresponding one decreased by approximately the same amount, leaving $\bar{V}_m$ and $\bar{V}_m$ largely unchanged. Thus, bends result in little if any scattering of the conduction electrons¹⁴ and small-diameter armchair nanotubes should behave as true electron waveguides.

The unconventional scaling rule proposed here predicts exceptional ballistic transport properties for single-walled weakly disordered armchair nanotubes. This prediction can be understood in physical terms as follows. In the weak scattering limit, electrons injected at ϵ_F enter ring states making up the a_1 and a_2 bands that are rotationally invariant under C_{N_B} rotations about the $[N_B, N_B]$ tubule axis. For small-diameter tubes, scattering of these conduction electrons by the disorder is only between ring states of this symmetry due to energy conservation. However, any linear combination of these ring states extends around the full circumference of the tube. Thus, the closest we can come to a classical picture of a nanotube conduction electron in the two-band basis is a doughnut-like wave packet confined along the tube but extended around its circumference. This 'doughnut' wave packet experiences an effective disorder that is an average of the real disorder over the tube's circumference ($\propto N_B$). The mean square amplitude of this effective disorder is then reduced relative to that of the real disorder by a factor of N_B whenever the real disorder fluctuates on the scale of a carbon–carbon bond length. Hence—so long as the two-band model remains valid—the larger the tube's diameter grows, the greater the reduction and the more freely the 'doughnut' wave packet propagates along the tube leading to an enhancement of l by a factor of N_B . For [10,10] nanotubes which have $N_B = 10$, this enhancement is large, affording the opportunity for directly observing ballistic electron transport over unprecedented distances for such laterally confined systems⁴. We do not expect that electron–electron interactions will alter this conclusion because spin-split states should still average the disorder over the circumference of the tube. □

Received 30 October 1997; accepted 23 February 1998.

- Mintmire, J. W., Dunlap, B. I. & White, C. T. Are fullerene tubules metallic? *Phys. Rev. Lett.* **68**, 631–634 (1992).
- Hamada, N., Sawada, A. & Oshiyama, A. New one-dimensional conductors: graphitic micro-tubules. *Phys. Rev. Lett.* **68**, 1579–1581 (1992).
- Saito, R., Fujita, M., Dresselhaus, G. & Dresselhaus, M. S. Electronic structure of chiral graphene tubules. *Appl. Phys. Lett.* **60**, 2204–2206 (1992).
- Tans, S. J. *et al.* Individual single-wall carbon nanotubes as quantum wires. *Nature* **386**, 474–477 (1997).
- Mott, N. F. & Twose, W. D. The theory of impurity conduction. *Adv. Phys.* **10**, 107–163 (1961).
- Tess, A. *et al.* Crystalline ropes of metallic carbon nanotubes. *Science* **273**, 483–487 (1996).
- Thouless, D. J. Maximum metallic resistance in this wires. *Phys. Rev. Lett.* **39**, 1167–1169 (1977).
- Anderson, P. W., Thouless, D. J., Abrahams, E. & Fisher, D. S. New method for a scaling theory of localization. *Phys. Rev. B* **22**, 3519–3526 (1980).
- Thouless, D. J. Localization distance and mean free path in one-dimensional disordered systems. *J. Phys. C* **6**, L46–L51 (1973).
- Todorov, T. N. Calculation of the residual resistivity of three-dimensional quantum wires. *Phys. Rev. B* **54**, 5801–5813 (1996).
- White, C. T., Robertson, D. H. & Mintmire, J. W. Helical and rotational symmetries of nanoscale graphitic tubules. *Phys. Rev. B* **47**, 5485–5488 (1993).
- Harrison, W. A. *Electronic Structure and Properties of Solids* 550 (Freeman, San Francisco, 1980).
- Ruoff, R. S. *et al.* Radial deformation of carbon nanotubes by van der Waals forces. *Nature* **364**, 514–516 (1993).
- Kane, C. L. & Mele, E. J. Size, shape, and low energy electronic structure of carbon nanotubes. *Phys. Rev. Lett.* **78**, 1932–1935 (1997).

Acknowledgements. C.T.W. thanks D. G. Pettifor for hospitality during his sabbatical stay at Oxford and N. J. Long and D. H. Robertson for technical assistance. T.N.T. thanks EPSRC and St John's College, Oxford for support. Support from the US Office of Naval Research and Hewlett Packard is also acknowledged.

Correspondence and requests for materials should be addressed to C.T.W. at NRL (e-mail: white@alchemy.nrl.navy.mil).

Intertwined symmetry of the magnetic modulation and the flux-line lattice in the superconducting state of TmNi₂B₂C

M. R. Eskildsen*, K. Harada†, P. L. Gammel‡, A. B. Abrahamsen*, N. H. Andersen*, G. Ernst‡, A. P. Ramirez‡, D. J. Bishop‡, K. Mortensen*, D. G. Naugle§, K. D. D. Rathnayaka§ & P. C. Canfield||

* Risø National Laboratory, PO Box 49, DK-4000 Roskilde, Denmark

† Advanced Research Laboratory, Hitachi Ltd, Hatoyama, Saitama 350-03, Japan

‡ Bell Laboratories, Lucent Technologies, 700 Mountain Avenue, Murray Hill, New Jersey 07974, USA

§ Physics Department, Texas A&M University, College Station, Texas 77843, USA

|| Ames Laboratory and Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011, USA

Materials that can in principle exhibit both superconductivity and ferromagnetism are caught in a dilemma: both states represent long-range order, but are in general mutually exclusive. When the material favours a ground state with a large magnetic moment, as is the case for Er₂Rh₄B (ref. 1), superconductivity is destroyed. For superconductivity to persist, the magnetic structure would need to adopt an antiferromagnetic modulation of short enough wavelength to ensure a small net moment on the length scale of the superconducting coherence length. The intermetallic borocarbide superconductors^{2–4} RNi₂B₂C (where R is a rare-earth element) have shed new light on this balance between magnetism and superconductivity. The response of these materials in the superconducting state to a magnetic field is dominated by the formation of a flux-line lattice—a regular array of quantized magnetic vortices whose symmetry and degree of order are easily modified and thus can be expected to interact with an underlying

magnetic modulation. In $\text{TmNi}_2\text{B}_2\text{C}$, superconductivity and antiferromagnetic modulated ordering coexist below 1.5 K (refs 5–7). Here we present the results of a small-angle neutron-scattering study of this compound which show that the structure of the magnetic modulation and the symmetry of the flux-line lattice are intimately coupled, resulting in a complex phase diagram.

Here we report on the results of simultaneous small-angle neutron-scattering (SANS) studies of the flux-line lattice (FLL) and magnetic order in $\text{TmNi}_2\text{B}_2\text{C}$ (Tm:1221), augmented with extensive characterization by transport and magnetization. Our single-crystal Tm:1221 sample was grown⁸ using a high temperature flux method. To increase neutron penetration, isotopically enriched ^{11}B was used. The single-crystal platelet ($4 \times 6 \times 1$ mm) had the c -axis in the thin direction. An overview of the phase diagram for the superconducting and magnetic states is shown in Fig. 1 inset. The material is superconducting for temperatures below $T_c = 11$ K. For fields applied parallel to the c -axis, as was the case throughout the experiment, the upper critical field saturates near $T = 5$ K and $H_{c2} = 10$ kOe due to the thulium (Tm) sublattice magnetization⁵.

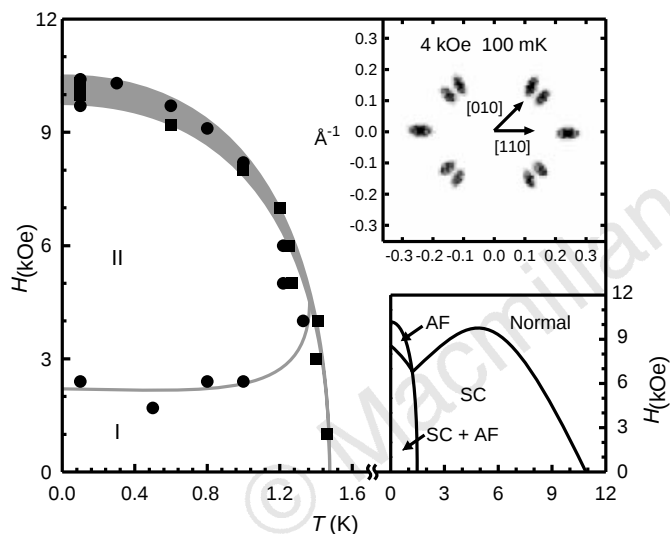


Figure 1 The magnetic phase diagram for $\text{TmNi}_2\text{B}_2\text{C}$. The zero-field magnetic modulated phase ordered along $[110]$ with $\mathbf{q}_{\text{mll}} = 0.094(\mathbf{a}^* + \mathbf{b}^*)$ is identified as I. At 2 kOe, there is a sharp magnetic transition into a complex modulated structure identified as II, with additional new split peaks appearing around the equivalent $[100]$ and $[010]$ directions. The intensity of the new reflections grow with the field until, above 4 kOe, all reflections are of roughly equal intensity. Above 4 kOe, the intensities of all reflections decrease slowly, vanishing at a second magnetic transition at 10 kOe at 100 mK. The squares show where the $\mathbf{q}_{\text{mll}}$ reflections vanish, and the circles show where the $\mathbf{q}_{\text{mll}}$ reflections vanish. Due to the slow field dependence there is an uncertainty on the upper transition field indicated by the shaded region. Upper inset shows a diffraction pattern obtained at $H = 4$ kOe and $T = 100$ mK using $3\text{-}\text{\AA}$ neutrons. The smearing of the peaks is an artefact of the coarse energy resolution ($\Delta\lambda_n/\lambda_n = 18\%$) on the SANS beamline. The missing peaks on the top and bottom are due to this image being an integration over a rock about the vertical axis, which does not satisfy the Bragg condition for those reflections. The crystal axes are shown by the arrows. The locations of the diffraction peaks, as well as the $15^\circ \pm 0.5^\circ$ splitting of the $\mathbf{q}_{\text{mll}}$ peaks, are only weakly temperature- and field-dependent. The magnitude of $\mathbf{q}_{\text{mll}}$ is 22% below that of $\mathbf{q}_{\text{mll}}$. The splitting of the peaks along $[100]$ and $[010]$ can be suppressed at high fields by cooling the sample in the saturated paramagnetic state ($H > 11$ kOe) and then lowering the field. Lower inset shows the combined phase diagram for superconductivity (SC) and magnetism (AF). The upper critical field H_{c2} is determined from a resistance onset criterion. The peak in H_{c2} near 5 K is associated with the large Tm local moment. As the Tm moment saturates below 4 K, the normal state resistance drops 25% due to reduced spin scattering. The regime of coexisting superconductivity and magnetism below 1.5 K is the focus of this Letter.

Below $T_N = 1.5$ K there is a magnetic transition into an incommensurate long-period (~ 25 Å) transverse-modulated antiferromagnetic state which coexists with the superconductivity^{6,7}. Initially this transition seems to have only a modest effect on the superconductivity, as H_{c2} shows only a gradual rise below T_N owing to the trade-off between the changes in the Tm susceptibility and the normal state mean free path.

We focus first on neutron-scattering studies of the magnetic order in an applied field. The experiments were performed using the SANS beamline in the cold neutron guide hall of the DR3 research reactor at Risø National Laboratory. The sample was mounted in the tailpiece of a dilution refrigerator equipped with single-crystal sapphire windows for neutron scattering. Temperatures in the range from 50 mK to 4 K were easily stabilized, and a superconducting magnet applied fields up to 13 kOe parallel to both the incident neutrons and the c -axis of the crystal. Bragg scattered neutrons were detected by an area detector in an evacuated chamber. For magnetic studies, a neutron wavelength of $3\text{ }\text{\AA}$ was used, and the detector was positioned 1.2 m from the sample. Scattering studies of magnetic order on the SANS beamline probe only a small part of reciprocal space, precluding detailed determination of magnetic spin configurations.

The magnetic diffraction data are summarized by the phase diagram in Fig. 1. The low-field region I is the incommensurate transversely modulated state previously studied only in zero field^{6,7}. The Tm spins order into a squared spin density wave with scattering vector $\mathbf{q}_{\text{mll}} = 0.094(\mathbf{a}^* + \mathbf{b}^*)$ and the moment parallel to the c -axis. A magnetic field parallel to the c -axis induces a nearly temperature-independent transition at 2 kOe into region II. Figure 1 inset shows a magnetic diffraction pattern obtained in region II. The new modulation, which coexists with the low-field state, is split about the $[100]$ direction with a magnitude $|\mathbf{q}_{\text{mll}}| = 0.21\text{ }\text{\AA}^{-1}$, 22% below

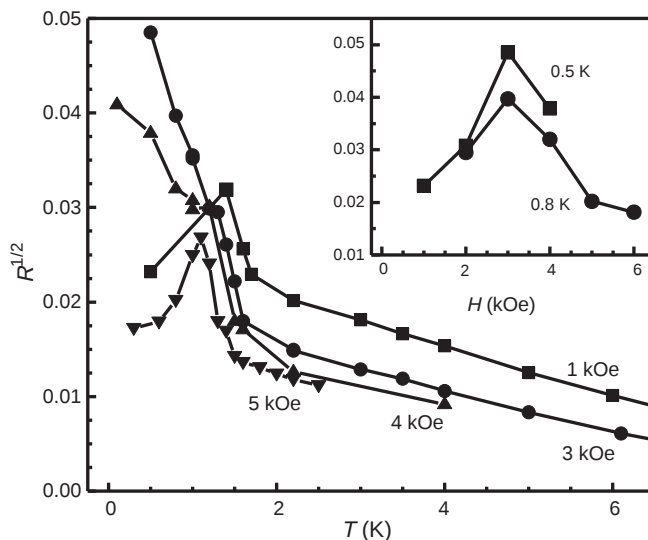


Figure 2 The temperature dependence of the integrated reflectivity R for the FLL at several different magnetic fields between 1 and 5 kOe. Above T_N , the reflectivity follows the simple form $\sqrt{R} \propto (T_c - T)$, which is the London limit of the form factor with the mean field form of the penetration depth, λ . The reflectivity is always sharply peaked at the magnetic transition, which could be a result of the reduction in the effective λ due to the susceptibility near the transition. In the field regime where the FLL shows two split square domains, the reflectivity increases dramatically at low temperatures, giving a peak in the reflectivity with field upon crossing this nearly temperature independent I–II magnetic transition as shown in the inset. The rocking-curve widths, over which the reflectivity is integrated, are not shown, but are smooth and featureless over this entire regime, consistent with a gradual disordering of the FLL with decreasing temperature throughout the superconducting state.

q_{ml} . The splitting is $15^\circ \pm 0.5^\circ$ and only weakly dependent on temperature and field. This splitting is not the result of a simple magneto-elastic distortion, as the q_{ml} peaks remain unsplit. The moment direction for this modulation is not fully constrained, but we speculate that it arises from components of the Tm spin, organized in two orthogonal domains. Between 2 and 4 kOe the reflectivity of the new peaks grows rapidly, while the reflectivity of the q_{ml} peaks drops correspondingly. Between 4 and 10 kOe all magnetic reflections have roughly equal, and falling intensity.

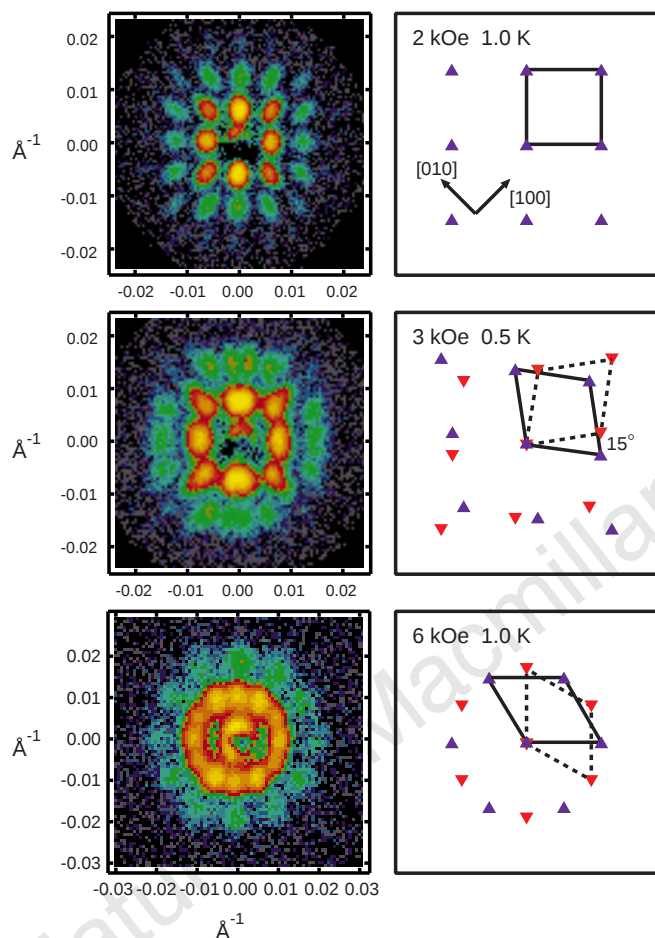


Figure 3 The different FLL structures observed in Tm:1221. The left-hand column shows the measured diffraction pattern, integrated over a rocking curve, and the right-hand column shows our real-space interpretation of the images. Where the FLL is composed of two equivalent domains, they are represented by up and down triangles. The crystal orientation is shown in the top panel and is identical for all fields. It is an inherent property of any two-dimensional structure that the real and reciprocal space configurations are identical except for a 90° rotation and a rescaling of the axes. In all cases the FLL is orientated by the [110] type crystal lattice directions, which are horizontal and vertical in both the real and reciprocal space pictures. Top row, diffraction pattern at $T = 1 \text{ K}$, $H = 2 \text{ kOe}$, showing a well ordered square lattice regime showing at least four orders of Bragg peaks. Middle row, diffraction pattern at $T = 0.5 \text{ K}$, $H = 3 \text{ kOe}$ above the first magnetic transition. There are two domains of rhombohedral FLL, as is best seen in the (2,0) Bragg peaks which are split by $15^\circ \pm 1^\circ$ around the [110] crystallographic direction. Bottom row, diffraction pattern at $T = 1.0 \text{ K}$, $H = 6 \text{ kOe}$. Here there are two domains of hexagonal FLL, orientated by the [110] crystal lattice directions. Several orders of diffraction peaks are visible, suggesting well ordered domains. The FLL undergoes a two-step structural symmetry transition from a square FLL below 3 kOe to a hexagonal FLL above 5 kOe. First the FLL undergoes a rhombic distortion along the [100] crystalline direction giving rise to two domains of distorted square FLLs with a splitting of 15° at 3 kOe. As the field is further increased, a hexagonal FLL, pinned to the [110] direction, is observed above 5 kOe.

Finally, at $H = 10 \text{ kOe}$ and $T = 100 \text{ mK}$, there is a second magnetic transition, presumably into a saturated paramagnetic state as we observe significant extra scattering near $q = 0$ for fields above H_{c2} . Unlike the sharp disappearance of the magnetic reflections with temperature in zero field^{6,7}, the intensity of the magnetic scattering shows a very slow field dependence, indicating critical scaling behaviour.

By simply changing the neutron wavelength to 5–11 $\text{\AA}$ and the sample-to-detector distance to 6 m, we can explore the Bragg reflections from the FLL. At temperatures above the Néel transition a square FLL orientated in the [110] direction was seen, in agreement with all previous studies of borocarbide superconductors^{9–11}. In this case, the field dependence of the FLL reflectivity and the Ginzburg–Landau expressions^{12,13} for the form factor were used to determine the penetration depth, λ , and coherence length, ξ , at $T = 1.9 \text{ K}$. We find $\lambda = 800 \text{ \AA}$ and $\xi = 150 \text{ \AA}$, in reasonable agreement with the literature⁵. The temperature dependence of the reflectivity, dominated by the penetration depth, agrees with the mean-field limit $\sqrt{R} \propto (T_c - T)$ as evident from Fig. 2. This temperature dependence is in agreement with previous observations⁹ in $\text{ErNi}_2\text{B}_2\text{C}$.

Both the symmetry of the FLL and its reflectivity show dramatic behaviour at and below T_N . At all fields, there is a sharp increase in the reflectivity on crossing the magnetic transition, as shown in Fig. 2. In addition, shown in the inset to Fig. 2, one finds that the reflectivity also has a peak on crossing the nearly temperature-independent transition between magnetic phases I and II. Clearly, it is no longer possible to use the simple Ginzburg–Landau expressions to model this effect, as they imply a monotonic decrease in reflectivity with both field and temperature. The sharp peaks may be a consequence of the reduction in λ due to the susceptibility correction^{14,15} near these transitions. Additional low-temperature magnetization data are needed to verify this picture.

More striking are the changes in the FLL symmetry in the vicinity of the magnetic transitions. Depending on field and temperature, we find three different FLL symmetries. Examples of the diffraction patterns observed and their real-space interpretations are shown in Fig. 3. In addition, Fig. 4 shows a composite phase diagram including the different FLL symmetries, the magnetic phases and

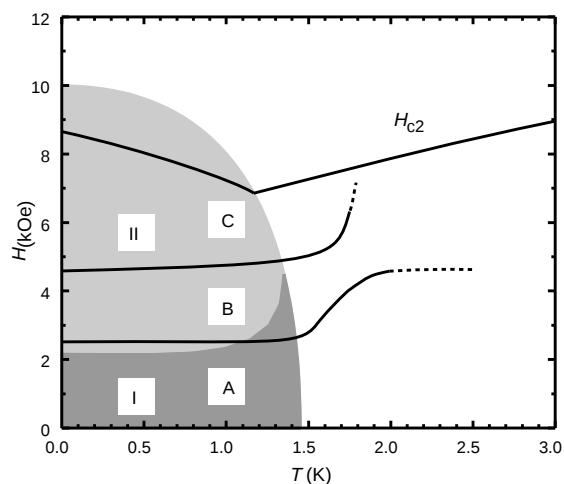


Figure 4 The low-temperature composite phase diagram, including bulk superconductivity, the FLL symmetry and magnetic phases in Tm:1221. The upper critical field H_{c2} is determined from transport. The magnetic phase diagram is from Fig. 1, and is shown by the shaded regions. The solid lines denote the boundaries between the different FLL symmetries, with dashed extensions where there is no experimental data. We note the strong correlation between the FLL symmetry and the magnetic order, especially near the 2 kOe transition. We denote the different FLL symmetries by A (square), B (rhombohedral) and C (hexagonal).

the upper critical field, H_{c2} . The square FLL, shown in the top row of Fig. 3, is stable below 2 kOe for all temperatures. We note that although the magnetic order below T_N is also square symmetric, other effects, including in-plane Fermi surface anisotropy¹⁶ and H_{c2} anisotropy¹⁷, can stabilize the square FLL. As the square FLL is stable in most of the superconducting phase outside the regime of magnetic ordering, this symmetry certainly cannot be solely attributed to the magnetic order. In the magnetically ordered state, coincident with the transition at 2 kOe, the FLL undergoes a rhombic distortion, as shown in the middle of row of Fig. 3. The two rhombohedral domains, rotated by 90°, are each distorted by $15^\circ \pm 1^\circ$, the same angular splitting as seen in the new magnetic modulation in this region. This symmetry remains stable between 3 and 4 kOe. Above 4 kOe, the FLL undergoes a second transition into two hexagonal domains, also rotated by 90°, shown in the bottom row of Fig. 3. This cannot be an extension of the low-field rhombohedral distortion, as that would result in hexagonal domains orientated in the [100] direction, rather than the observed [110] as is evident from Fig. 3. Instead, this change in orientation implies a discontinuous transition. Above 4 kOe, all 12 magnetic reflections are of equal intensity, giving the same number of reflections as the FLL, albeit with a symmetry distorted from pure hexagonal. The origin of the FLL symmetry transitions in relation to magnetic transition is still an open question, as we cannot from the present data determine if one drives the other or *vice versa*.

It is important to point out the difference between the symmetry transitions reported here, and our previous reports of the low-field transition ($H < 1$ kOe) in $\text{ErNi}_2\text{B}_2\text{C}$ (ref. 10), and $\text{YNi}_2\text{B}_2\text{C}$ and $\text{LuNi}_2\text{B}_2\text{C}$ (ref. 11) where the transitions proceed continuously through a rhombic distortion and is understood from anisotropic flux line interactions due to the high in-plane anisotropy of the Fermi surface¹⁶. The low-field region was not probed in the SANS measurements, but Bitter decorations at 50 Oe and 4.2 K verified a hexagonal FLL. Previous studies of the interaction between magnetism and the superconducting FLL in $\text{ErNi}_2\text{B}_2\text{C}$ showed a rotation due to the internal field direction and a disordering due to increased pinning⁹. We believe that these effects, while convincing evidence of the strong coupling between the two types of order, can be understood within the scattering time approximation. The observations reported in here will require a deeper and more subtle understanding.

The splitting of the magnetic peaks with wavevector $\mathbf{q}_{\text{mll}}$ can be suppressed between 4 and 10 kOe by cooling the sample in the saturated paramagnetic state and then reducing the field into the hexagonal FLL state. Such low-temperature hysteresis is common in magnetic systems. In $\text{HoNi}_2\text{B}_2\text{C}$, hysteresis in the metamagnetic transitions has been reported¹⁸, and in $\text{DyNi}_2\text{B}_2\text{C}$ the associated residual moment has been seen to suppress superconductivity¹⁹. This hysteresis is not seen in the FLL. If the FLL is the determining factor in this problem, then this hysteresis in the split of the magnetic modulation could be ascribed to passing through a FLL symmetry transition. Further studies into the details of the symmetries and the transitions between them are needed to identify the driving mechanisms, and to shed more light on the intimate connection between superconductivity and magnetism in this material. □

Received 31 December 1997; accepted 23 March 1998.

- Crabtree, G. W. *et al.* Anisotropic superconducting and magnetic properties of a single crystal of ErRh_4B_4 . *Phys. Rev. Lett.* **49**, 1342–1345 (1982).
- Siegrist, T. *et al.* The crystal structure of superconducting $\text{LuNi}_2\text{B}_2\text{C}$ and the related phase LuNiBC . *Nature* **367**, 254–256 (1994).
- Cava, R. J. *et al.* Superconductivity in the quaternary intermetallic compounds $\text{LnNi}_2\text{B}_2\text{C}$. *Nature* **367**, 252–253 (1994).
- Nagarajan, R. *et al.* Bulk superconductivity at an elevated temperature ($T_c \approx 12$ K) in a nickel containing alloy system Y-Ni-B-C . *Phys. Rev. Lett.* **72**, 274–277 (1994).
- Cho, B. K. *et al.* Magnetic and superconducting properties of single-crystal $\text{TMNi}_2\text{B}_2\text{C}$. *Phys. Rev. B* **52**, 3676–3683 (1995).
- Lynn, J. W. *et al.* Magnetic order and crystal structure in the superconducting RNiB_2C materials. *Phys. Rev. B* **55**, 6584–6598 (1997).

- Sternlieb, B. *et al.* Single crystal diffraction study of the magnetic structure of $\text{TMNi}_2\text{B}_2\text{C}$. *J. Appl. Phys.* **81**, 4937–4939 (1997).
- Cho, B. K. *et al.* Magnetism and superconductivity in single crystal $\text{ErNi}_2\text{B}_2\text{C}$. *Phys. Rev. B* **52**, 3684–3695 (1995).
- Yaron, U. *et al.* Microscopic coexistence of magnetism and superconductivity in $\text{ErNi}_2\text{B}_2\text{C}$. *Nature* **382**, 236–238 (1996).
- Eskildsen, M. R. *et al.* Observation of a field-driven structural phase transition in the flux line lattice in $\text{ErNi}_2\text{B}_2\text{C}$. *Phys. Rev. Lett.* **78**, 1968–1971 (1997).
- Eskildsen, M. R. *et al.* Structural stability of the square flux line lattice in $\text{YNi}_2\text{B}_2\text{C}$ and $\text{LuNi}_2\text{B}_2\text{C}$ studied with small angle neutron scattering. *Phys. Rev. Lett.* **79**, 487–490 (1997).
- Gammel, P. L. *et al.* Small angle neutron scattering study of the magnetic flux-line lattice in single crystal 2H-NbSe_2 . *Phys. Rev. Lett.* **72**, 278–282 (1994).
- Kleiman, R. N. *et al.* Neutron diffraction from the vortex lattice in the heavy-fermion superconductor UPt_3 . *Phys. Rev. Lett.* **69**, 3120–3124 (1992).
- Gray, K. E. Ginzburg-Landau equations, interphase surface energy, and the intermediate state of superconductors with a paramagnetic normal state. *Phys. Rev. B* **27**, 4157–4160 (1983).
- Tachiki, M., Matsumoto, H. & Umezawa, H. Mixed state in magnetic superconductors. *Phys. Rev. B* **20**, 1915–1927 (1979).
- Kogan, V. G. *et al.* Vortex lattice transitions in borocarbides. *Phys. Rev. B* **55**, R8693–R8696 (1997).
- De Wilde, Y. *et al.* Scanning tunneling microscopy observation of a square Abrikosov lattice in $\text{LuNi}_2\text{B}_2\text{C}$. *Phys. Rev. Lett.* **78**, 4273–4277 (1997).
- Canfield, P. C. Angular dependence of metamagnetic transitions in $\text{HoNi}_2\text{B}_2\text{C}$. *Phys. Rev. B* **55**, 970–976 (1997).
- Peng, Z. Q., Krug, K. & Winzer, K. Large hysteresis effect and reentrant behavior in $\text{DyNi}_2\text{B}_2\text{C}$ at temperatures $T < 2$ K. *Phys. Rev. Lett.* (submitted).

Acknowledgements. We thank C. Stassis, P. Hedegård and J. Jensen for discussions. This work was supported by NATO. M.R.E. is supported by the Danish Research Academy, D.G.N. is supported by the Robert A. Welch Foundation and P.C.C. is supported by the Director of Energy Research, Office of Basic Energy Science.

Correspondence and requests for materials should be addressed to M.R.E. (e-mail: morten.eskildsen@risoe.dk).

Simulated response of the ocean carbon cycle to anthropogenic climate warming

Jorge L. Sarmiento*, Tertia M. C. Hughes*, Ronald J. Stouffer† & Syukuro Manabe††

* Program in Atmospheric and Oceanic Sciences, Princeton University, PO Box CN710, Princeton, New Jersey 08544, USA

† Geophysical Fluid Dynamics Laboratory/NOAA Princeton University, PO Box 308, Princeton, New Jersey 08542, USA

A 1995 report¹ of the Intergovernmental Panel on Climate Change provides a set of illustrative anthropogenic CO_2 emission models leading to stabilization of atmospheric CO_2 concentrations ranging from 350 to 1,000 p.p.m. (refs 1–4). Ocean carbon-cycle models used in calculating these scenarios assume that oceanic circulation and biology remain unchanged through time. Here we examine the importance of this assumption by using a coupled atmosphere–ocean model of global warming⁵ for the period 1765 to 2065. We find a large potential modification to the ocean carbon sink in a vast region of the Southern Ocean where increased rainfall leads to surface freshening and increased stratification⁶. The increased stratification reduces the downward flux of carbon and the loss of heat to the atmosphere, both of which decrease the oceanic uptake of anthropogenic CO_2 relative to a constant-climate control scenario. Changes in the formation, transport and cycling of biological material may counteract the reduced uptake, but the response of the biological community to the climate change is difficult to predict on present understanding. Our simulation suggests that such physical and biological changes might already be occurring, and that they could substantially affect the ocean carbon sink over the next few decades.

We use the GFDL atmosphere–ocean climate model⁷ to which we

† Present address: Earth Frontier Research System, 7th Floor, Seavans Building-N, 1-2-1 Shibaura, Minato-ku, Tokyo 105, Japan.

have coupled an ocean carbon model⁸. A constant climate 'control' model fixes CO₂ and other long-lived greenhouse gases at their initial value. The global warming ('GW') model is that of Haywood *et al.*⁵, which includes estimates of the radiative forcing based on observed CO₂ (Fig. 1a) and other long-lived greenhouse gases from 1765 to 1990, and predicted future increases of these gases until 2065 (see Methods). Also included is an estimate of the direct effect of sulphate aerosols. The model prediction of warming in global mean surface air temperatures is in reasonable agreement with observations⁵. Despite limitations in the model, and our inability to know whether the climate sensitivity of this model is correct, the fact that the model reproduces the observed temperature record encourages us to use it as a framework for examining the response of the global carbon cycle to climate change in the next century.

The ocean carbon component of the coupled atmosphere–ocean model solves the full carbon system equations and includes the 'biological pump' (in which organic matter and CaCO₃ are formed at the ocean surface, then sink and are released at depth) parametrized as in the previous century-scale coupled model study⁸ of Sarmiento and Le Quéré (Fig. 1b; see Methods). Given our present understanding, any model of how biology will respond to global warming must be considered highly speculative. Our approach was to run a simple model based on assuming a minimum change in the production and export of organic matter and CaCO₃. This 'constant-biota' model fixes the production and export at their initial values except where the major nutrient phosphate becomes depleted at the surface, in which case production is set to zero. Such a fixed production model might be expected if whatever limits growth in the present ocean (for example, iron supply by windborne dust⁹) were to remain constant. We fix the time history of atmospheric CO₂ as in Fig. 1a and allow the CO₂ to invade the ocean. One simulation was performed in the control climate model to provide a 'baseline' with constant climate. This is similar to the assumption of constant ocean circulation (except that climate variability is included) that was made in the Intergovernmental Panel on Climate Change (IPCC) CO₂ stabilization models^{1–4}. A second simulation was performed in the GW model.

We find that the constant-biota GW model takes up almost as much CO₂ as the constant-biota baseline model (Fig. 1b; rows A and E of Table 1). The 1980–89 mean annual uptake is in the middle of the IPCC estimate of 2.0 ± 0.8 Pg C yr⁻¹ for ocean models without global warming¹. However, an analysis of the processes contributing to the global warming simulation reveals that, without the biological pump, there would be a very large negative influence on the oceanic uptake owing to warming of the ocean and changes in transport (Fig. 1c; rows B and C of Table 1). The solubility GW model that we use to isolate these processes (see Methods) shows a divergence from its baseline beginning before ~1970 (Fig. 1c). By 1990 the cumulative solubility GW uptake is 27 Pg C lower than its baseline, and by 2065 the decrease is 124 Pg C (rows B and C of Table 1). The latter is about half of the total fossil CO₂ emissions from the beginning of the industrial revolution until today. The constant-biota model counteracts these decreases, but we need to know whether other scenarios would do the same.

As a reasonable lower limit to the impact of the biological pump, we examined a 'constant phosphate' model that fixes the surface phosphate content through time, such as might be expected if the primary limit to phytoplankton growth were a limiting micronutrient that is supplied primarily from within the ocean (for example, possibly dissolved iron supplied by deep waters of the present Southern Ocean¹⁰). The 1990 to 2065 cumulative uptake in the constant phosphate model (Fig. 1b) is 219 Pg C. This is 79 Pg C less than its control run, as contrasted with a difference of only 27 Pg C for the constant-biota model. An upper-limit model of complete phosphate consumption (a 'super-biotic' model similar to the iron-fertilization scenario explored by Sarmiento and Orr¹¹ and to one of the proposed hypotheses for the lowering of CO₂ during the last ice age^{12–14}) shows a 1990 to 2065 cumulative uptake of 409 Pg C. Although such an extreme scenario would never be realized, it is clear, from the range between the lower limit of 219 Pg C and the super-biotic uptake of 409 Pg C, that changes in biology could easily modify the ocean carbon sink substantially between now and the middle of the twenty-first century.

An important result of the simulations is that most of the impact

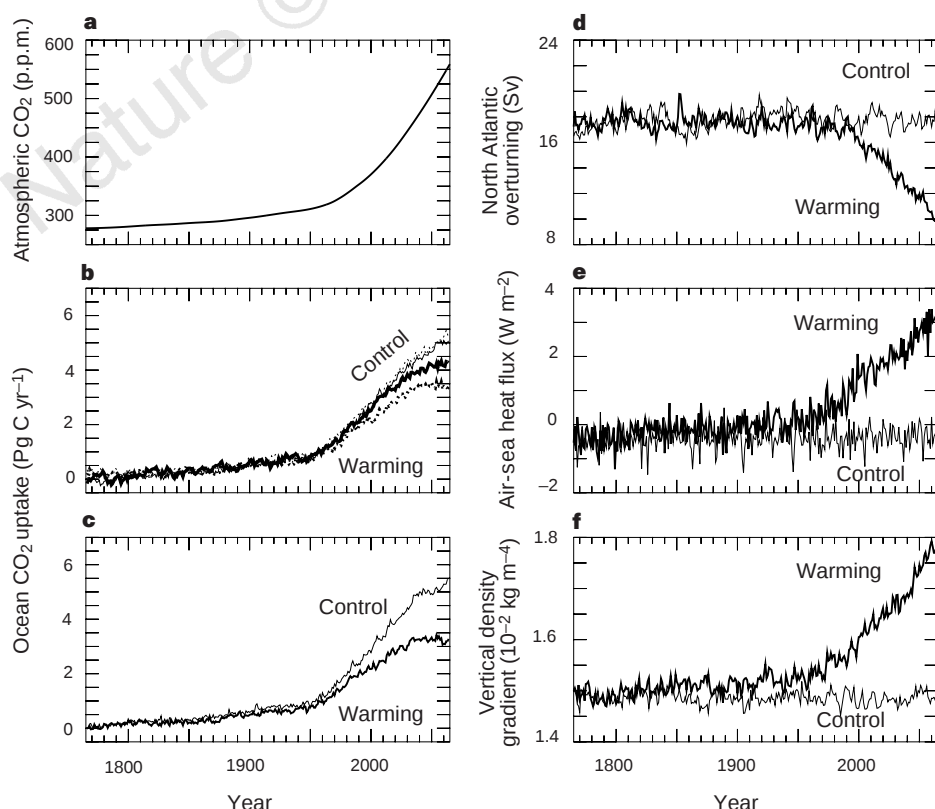


Figure 1 Time series of model boundary conditions and predictions. **a**, Atmospheric CO₂ based on observations before 1990 and the IPCC IS92a model²³ thereafter. **b**, Annual ocean uptake of carbon by the biology model. Four simulations are shown: the constant-biota control model (thin solid line), the constant-biota GW model (thick solid line), the constant-phosphate control model (thin dotted line) and the constant-phosphate GW model (thick dotted line). **c**, Annual ocean uptake of carbon by the 'solubility' model. **d**, Maximum value of overturning in the thermohaline cell of the North Atlantic in units of Sverdrups (1 Sv = 10^6 m³ s⁻¹). **e**, Global mean of the air-sea heat flux. **f**, Global mean vertical density gradient at the base of the first layer of the model (50.9 m) for latitudes polewards of 30° in both hemispheres.

Table 1 Oceanic uptake of anthropogenic CO₂ in constant-biota model

	Annual uptake (Pg C yr ⁻¹)		Cumulative uptake (Pg C)	
	1980–89	2056–65	1765–1990	1990–2065
A. Constant-biota baseline	1.96	5.03	112	289
B. Change due to variation in transport processes in absence of biological pump*	–0.17	–1.54	–5	–63
C. Change due to warming in absence of biological pump†	–0.20	–0.50	–23	–33
D. Change due to biological pump‡	+0.40	+1.26	+39	+69
E. Global warming constant-biota model	1.99	4.25	123	262

* The change due to variation in transport processes is calculated from the difference between the solubility baseline model and a solubility 'constant-temperature' GW scenario in which surface temperature was kept constant for calculation of surface carbon chemistry. An analysis of the transport out of the surface layer shows that the processes that decrease are vertical mixing and convective overturning. In contrast, advective transport increases. During the final decade of the simulation, the ratio of the decrease in convection to mixing is 0.24:0.76. The change in vertical mixing is mostly (>90%) accounted for by the isopycnal component, which is decreased owing to the change in the slope of the isopycnals. The advective contribution increases by an amount equal to 38% of the decrease in convective transport.

† The change due to warming is calculated by taking the difference between the solubility GW and solubility constant-temperature GW simulations.

‡ The biological contribution is estimated by taking the difference between the global warming constant-biota result and the constant biota baseline plus the change due to warming and the change in ocean transport processes in the solubility model; that is, $D = E - (A + B + C)$.

of warming and transport processes (as well as the compensatory effect of biology) occurs in the Southern Ocean (Fig 2a and b). A primary reason for the importance of the Southern Ocean is its dominance in the global oceanic uptake of anthropogenic carbon¹⁵. Similar changes that occur in the high latitudes of the North Atlantic are larger by unit area but smaller in total impact owing to the relatively small total area that is involved.

Because of the importance of the high latitudes in the carbon cycle response to global warming, it is tempting to ascribe the observed effects to changes in deep-water formation, as was suggested for an idealized century-scale study with the same model as ours⁸. In support of this, our simulations show a marked decrease in the circulation of North Atlantic Deep Water beginning as early as the next decade (Fig. 1d). However, in the period between now and the middle of the twenty-first century, this and other changes in vertical advection have very little impact on the oceanic uptake of carbon from the atmosphere. Indeed, the region of the Southern Ocean where the maximum changes in air–sea flux occur is characterized by massive upwelling due to Ekman divergence. This upwelling actually increases slightly in response to the global climate warming.

An analysis of the term balances in the upper layer of the model

shows that the decrease in uptake in the solubility GW model is due primarily to the net flux of heat into the ocean (perturbation heat flux) that results from climate warming (which decreases CO₂ solubility) (Table 1; Fig. 1e and 2c), and a decrease in the magnitude of the vertical component of isopycnal mixing and convective overturning (Table 1; Fig. 2d). Note that the correct diagnostic for the effect of heating on the air–sea CO₂ flux is the perturbation of the atmosphere–ocean heat flux, which measures the present rate at which the ocean is being heated. The increase in temperature that results from this input of heat (which is concentrated at lower latitudes than the perturbation of the heat flux) is a diagnostic for the integrated effect of heating, not of its instantaneous rate. The heat flux perturbation effect predominates in the early period, accounting for 82% of the decrease in the cumulative uptake from 1765 to 1990 (Table 1). However, between 1990 and 2065, transport processes (mixing, convective overturning and advection) account for 66% of the decrease in cumulative uptake; during the final decade, the contribution of transport processes has increased to 75% of the decrease.

The high-latitude heat flux perturbation and changes in ocean transport processes are a consequence of increased stratification of the upper water column (Fig. 1f). The increased stratification causes

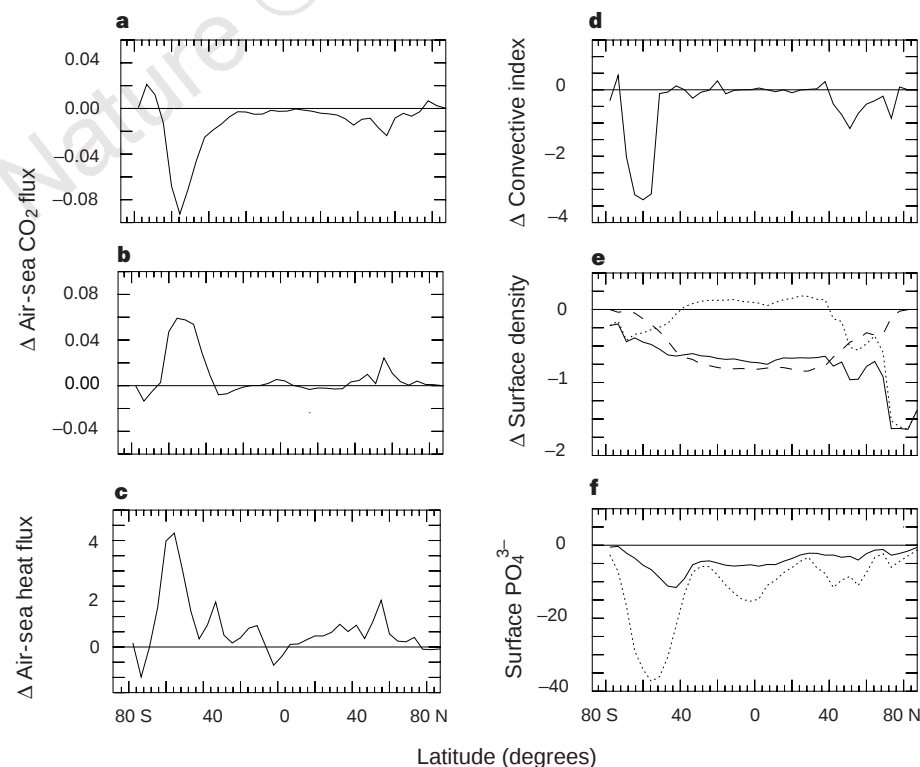


Figure 2 Zonal integrals for the decade 2056–65 of: **a**, the change in air–sea CO₂ fluxes (positive for ocean uptake) due to warming and changes in ocean transport estimated by the difference between the solubility GW and the solubility baseline (Pg degree⁻¹ yr⁻¹); **b**, the change in air–sea CO₂ fluxes due to the biological pump estimated by taking the difference between the constant-biota GW and the constant-biota baseline air–sea CO₂ fluxes, after removing the changes due to warming and transport (Pg degree⁻¹ yr⁻¹); **c**, GW minus baseline heat fluxes in W m⁻²; and **d**, GW minus baseline of the convection index (10⁶ m times the fraction of the year during which convection occurs at the base of the first model layer). **e**, Zonal mean change in surface density (GW minus control model in units of kg m⁻³). The solid line is the total change in density. The dashed and dotted lines are the density changes due to temperature and salinity, respectively. **f**, Zonal integral (solid line) for the decade 2056–65 of constant biota GW minus constant biota baseline surface phosphate concentration in mmol m⁻². Also shown (dotted line) is the negative of the surface phosphate concentration in the baseline scenario. Note that the phosphate depletion is substantially less than the phosphate available in almost all latitude bands.

the decrease in vertical mixing along isopycnals (which are now more horizontal) as well as the decrease in convective overturning (Fig. 2d). These changes in transport decrease the vertical transport of carbon. They also decrease the upward mixing of warm waters from below. Because of this, surface temperatures do not increase as much as elsewhere, and even cool in some regions¹⁶, with the consequence that the perturbation of the heat flux is greatest in the Southern Ocean. The increased stratification also causes a gradual collapse of the thermohaline circulation⁶ in the entire deep ocean, which has a major role in the carbon balance on century timescales⁸.

Increased stratification at the surface is a global phenomenon resulting from increased temperature in low latitudes and freshening in high latitudes (Fig. 2e). In this connection we note that an earlier ocean-only study of the carbon cycle response to global warming over the industrial period was forced with warming but did not include changes in hydrology¹⁷. In our coupled model, the freshening of the high latitudes by increased precipitation is crucial to determining the response of the ocean carbon cycle to global warming⁶. The anthropogenic carbon uptake predicted by the previous study and our constant biota model are similar over the next 50–60 years, but the mechanisms driving that response are different.

The enhancement of atmospheric CO₂ uptake by the biological pump in the constant-biota simulation also occurs primarily in the high latitudes of the Southern Ocean (Fig. 2b), mirroring the decrease in CO₂ uptake in the solubility GW model (Fig. 2a). The dominance of the Southern Ocean is again due to its sensitivity to the changes in stratification and to the fact that this is the region of the world in which deep mixing is normally able to reach into the vast volume of deep water that holds excess biogenic carbon¹¹. The effect of the biological model on the air–sea flux of CO₂ depends on how the biological pump responds to the increased stratification and changes in ocean transport. We initialize the biological model with the downward biological pump flux in balance with the upward transport of excess dissolved inorganic carbon and nutrients; that is, there is no initial net flux of CO₂ across the air–sea interface. The slowing of vertical exchange in the GW model breaks this balance by decreasing the upward supply of excess dissolved inorganic carbon. The continued downward flux of organic carbon and CaCO₃ in the constant-biota GW model leads to an accumulation of excess biogenic dissolved inorganic carbon in the deep ocean. A diagnostic of the effect of decreased upward supply of deep carbon and nutrients in the constant-biota simulation is the decrease in surface phosphate concentration (Fig. 2f). The phosphate that is removed from the surface is accompanied by carbon. The resulting decrease in surface dissolved inorganic carbon leads to an increase in anthropogenic CO₂ uptake. The constant-phosphate model has a much smaller effect on the surface dissolved inorganic carbon concentration because it does not decrease the surface phosphate concentration.

An important caveat regarding our results is that all the ocean transport processes that change are parametrizations of phenomena that are not resolved by the model. It is urgent that simulations such as ours be attempted with models with different parametrizations of mixing and convection¹⁸. If similar results are obtained, we would urge consideration of a major international effort to monitor the Southern Ocean for changes in temperature, salinity and vertical stratification, as well as changes in the biological pump; and to study the processes that determine the response of the Southern Ocean to global warming. Our understanding of biological processes, in particular, is still rudimentary¹⁹.

Detection of changes in the carbon balance due to the biological pump should be possible, although challenging. Atmospheric oxygen is a possible tracer. Between 1995 and 2065, we infer an increase of 283 p.p.m. in the ratio of atmospheric oxygen to nitrogen in the constant-biota model. This signal (equivalent to 59 p.p.m. CO₂, with the addition of a carbon atom²⁰) would be readily detectable^{20,21}, but would have to be isolated from the signals from

fossil fuel burning and terrestrial biota²⁰. The contribution of oceanic warming to the oxygen signal could possibly be isolated from other processes by using changes in the atmospheric concentration ratio of noble gases to nitrogen (which differ in the temperature sensitivity of their solubilities; R. Keeling, personal communication). Another possible tracer of change is oceanic oxygen, whose global mean decreases by -8 mmol m^{-3} between 1990 and 2065. Locally, at a depth of $\sim 1,600 \text{ m}$ in the Southern Ocean, large regions southeast of the Pacific and south of the Indian Ocean show a decrease of more than -20 mmol m^{-3} , with troughs in excess of -50 mmol m^{-3} . These changes would be easy to measure, but present models are not sufficiently reliable to indicate where those regions would be. □

Methods

The GW climate model includes radiative forcing from CO₂ plus the effect of other long-lived greenhouse gases represented as their equivalent in CO₂. The estimated direct effect of sulphate aerosols is represented as an increase in surface reflectance. The thermal forcing is similar to that of Mitchell *et al.*²², with an increase of $\sim 1\%$ per year after 1990, which is similar to the IPCC IS92a model²³. The GW model does not include changes in radiative forcing by ozone or soot, changes in solar forcing, or the indirect effect of aerosols. The atmosphere and ocean models have approximately 4° horizontal resolution.

The biological pump consists of: (1) the formation of organic matter and CaCO₃ at the surface of the ocean; (2) export to depth by sinking of particulate organic matter and CaCO₃, as well as transport of dissolved organic matter; and (3) metabolism of organic matter and dissolution of CaCO₃ in the deep ocean to form 'excess' dissolved inorganic carbon and nutrients. The initial magnitude of the biological pump is determined before coupling the ocean model to the atmospheric model by requiring that the model fit the observed surface distribution of phosphate^{8,24}. The ratio of organic carbon to phosphorus is fixed at 120:1, with half the production going into particulate organic matter and half into dissolved organic carbon. Particulate organic matter is metabolized following a scaling depth obtained from sediment trap observations²⁵. Dissolved organic matter is metabolized as a first-order decay process. The ratio of CaCO₃ to phosphorus production and dissolution is determined at each depth level of the model by forcing the horizontal mean alkalinity profile predicted by the model towards observations. The CaCO₃-to-phosphorus ratio is fixed after coupling the ocean model to the atmosphere model. We ignore the small sediment burial term of both organic matter and CaCO₃.

Before coupling the ocean carbon model to the atmosphere–ocean climate model, we determine the total carbon content by fixing atmospheric CO₂ at $277.95 \mu\text{atm}$ and allowing it to invade the ocean-only model until the air–sea flux goes to zero. After the ocean carbon model has been coupled to the climate model, the coupled model is run for an additional 250 years with a fixed atmospheric CO₂ concentration before beginning the global warming simulations. The model is then run from 1765 to 2065 with atmospheric CO₂ increasing as in Fig. 1a and invading the ocean by gas exchange with a windspeed-dependent gas-exchange coefficient²⁶. In our predictions of anthropogenic CO₂ uptake, we ignore sediment dissolution of CaCO₃ and changes in calcification rates that would buffer the oceanic carbon chemistry. An important aspect of future increases in atmospheric CO₂ is that they will occur too rapidly for the sediment dissolution buffering mechanism to have a role until much later²⁷.

The solubility model is identical to the biological simulation except that the biological pump is removed. An additional change is to decrease the total amount of alkalinity in the ocean to the salinity-normalized mean of surface observations ($2,379 \text{ meq m}^{-3}$ at a salinity of 35‰, as contrasted with the global mean of $2,477 \text{ meq m}^{-3}$ at a salinity of 35‰ used in the biological model). This is done so that the surface carbon buffer capacity in the solubility model is comparable to that in the biological model. The solubility model permits us to examine the effect of warming and changes in ocean circulation without the biological response.

Received 7 November 1997; accepted 13 March 1998.

- Schimel, D. *et al.* in *Climate Change 1995* (ed. Houghton, J. T.) 76–86 (Cambridge Univ. Press, 1996).
- Schimel, D. *et al.* in *Climate Change 1994* (ed. Houghton, J. T.) 35–71 (Cambridge Univ. Press, 1995).

3. Enting, I. G., Wigley, T. M. L. & Heimann, M. *Future Emissions and Concentrations of Carbon Dioxide: Key Ocean/Atmosphere/Land Analyses* (CSIRO Division of Atmospheric Research, 31, 1994).
4. Sarmiento, J. L., Le Quéré, C. & Pacala, S. W. Limiting future atmospheric carbon dioxide. *Glob. Biogeochem. Cycles* **9**, 121–138 (1995).
5. Haywood, J. M., Stouffer, R. J., Wetherald, R. T., Manabe, S. & Ramaswamy, V. Transient response of a coupled model to estimate changes in greenhouse gas and sulfate concentrations. *Geophys. Res. Lett.* **24**, 1335–1338 (1997).
6. Manabe, S. & Stouffer, R. J. Century-scale effects of increased atmospheric CO₂ on the ocean–atmosphere system. *Nature* **364**, 215–218 (1993).
7. Manabe, S., Stouffer, R. J., Spelman, M. J. & Bryan, K. Transient responses of a coupled ocean–atmosphere model to gradual changes of atmospheric CO₂. Part 1: annual mean response. *J. Clim.* **7**, 85–118 (1994).
8. Sarmiento, J. L. & Le Quéré, C. Oceanic carbon dioxide uptake in a model of century-scale global warming. *Science* **274**, 1346–1350 (1996).
9. Coale, K. H. *et al.* A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean. *Nature* **383**, 495–501 (1996).
10. Johnson, K. S., Gordon, R. M. & Coale, K. H. What controls dissolved iron concentrations in the world ocean? *Mar. Chem.* **57**, 137–161 (1997).
11. Sarmiento, J. L. & Orr, J. C. Three-dimensional simulations of the impact of Southern Ocean nutrient depletion on atmospheric CO₂ and ocean chemistry. *Limnol. Oceanogr.* **36**, 1928–1950 (1991).
12. Siegenthaler, U. & Wenk, T. Rapid atmospheric CO₂ variations and ocean circulation. *Nature* **308**, 624–626 (1984).
13. Sarmiento, J. L. & Toggweiler, J. R. A new model for the role of the oceans in determining atmospheric pCO₂. *Nature* **308**, 621–624 (1984).
14. Knox, F. & McElroy, M. Changes in atmospheric CO₂, influence of marine biota at high latitudes. *J. Geophys. Res.* **89**, 4629–4637 (1984).
15. Sarmiento, J. L., Orr, J. C. & Siegenthaler, U. A perturbation simulation of CO₂ uptake in an ocean general circulation model. *J. Geophys. Res.* **97**, 3621–3646 (1992).
16. Manabe, S. & Stouffer, R. J. Multiple century response of a coupled ocean–atmosphere model to an increase of atmospheric carbon dioxide. *J. Clim.* **7**, 5–23 (1994).
17. Maier-Reimer, E., Mikolajewicz, U. & Winguth, A. Future ocean uptake of CO₂: interaction between ocean circulation and biology. *Clim. Dyn.* **12**, 711–721 (1996).
18. Gent, P. R., Willebrand, J., McDougall, T. J. & McWilliams, J. C. Parameterizing eddy-induced tracer transports in ocean circulation models. *J. Phys. Oceanogr.* **25**, 463–474 (1995).
19. Denman, K., Hofmann, E. & Marchant, H. In *Climate Change 1995* (ed. Houghton, J. T.) 483–516 (Cambridge Univ. Press, 1996).
20. Keeling, R. F., Najjar, R. P., Bender, M. L. & Tans, P. P. What atmospheric oxygen measurements can tell us about the global carbon cycle. *Glob. Biogeochem. Cycles* **7**, 37–68 (1993).
21. Bender, M., Ellis, T., Tans, P., Francey, R. & Lowe, D. Variability in the O₂/N₂ ratio of southern hemisphere air, 1991–1994: implications for the carbon cycle. *Glob. Biogeochem. Cycles* **10**, 9–22 (1996).
22. Mitchell, J. F. B., Johns, T. C., Gregory, J. M. & Tett, S. F. B. Climate response to increasing levels of greenhouse gases and sulphate aerosols. *Nature* **376**, 501–504 (1995).
23. Houghton, J. T., Callander, B. A. & Varney, S. K. (eds) *Climate Change 1992* (Cambridge Univ. Press, 1992).
24. Sarmiento, J. L., Murnane, R. & Quéré, C. L. Air–sea CO₂ transfer and the carbon budget budget of the North Atlantic. *Phil. Trans. R. Soc. Lond. B* **348**, 211–218 (1995).
25. Martin, J. H., Knauer, G. A., Karl, D. M. & Broenkow, W. W. VERTEX: carbon cycling in the northeast Pacific. *Deep-Sea Res.* **34**, 267–285 (1987).
26. Wanninkhof, R. Relationship between wind speed and gas exchange over the ocean. *J. Geophys. Res.* **97**, 7373–7383 (1992).
27. Archer, D., Kheshti, H. & Maier-Reimer, E. Multiple timescales for the neutralization of fossil fuel CO₂. *Geophys. Res. Lett.* **24**, 405–408 (1997).

Acknowledgements. We thank Corinne Le Quéré for adding the carbon component of the coupled model, Fortunat Joos for providing the atmospheric CO₂ figures, and Jerry Mahlman and Yasuhiro Yamamoto, as well as Klaus Keller, Francois Morel, Philippe Tortell and Ernst Maier-Reimer, for comments. The contributions of J.L.S. and T.M.C.H. were supported by the Office of Global Programs of the National Oceanic and Atmospheric Administration, and by the National Science Foundation. The paper was written in part while J.L.S. was visiting the Bermuda Biological Station for Research with support from EXXON Corporation; J.L.S. thanks Tony Knap for making the arrangements.

Correspondence and requests for materials should be addressed to J.L.S. (e-mail: jls@splash.princeton.edu).

Dynamic responses of terrestrial ecosystem carbon cycling to global climate change

Mingkui Cao* & F. Ian Woodward

Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

Terrestrial ecosystems and the climate system are closely coupled, particularly by cycling of carbon between vegetation, soils and the atmosphere. It has been suggested^{1,2} that changes in climate and in atmospheric carbon dioxide concentrations have modified the carbon cycle so as to render terrestrial ecosystems as substantial

carbon sinks^{3,4}; but direct evidence for this is very limited^{5,6}. Changes in ecosystem carbon stocks caused by shifts between stable climate states have been evaluated^{7,8}, but the dynamic responses of ecosystem carbon fluxes to transient climate changes are still poorly understood. Here we use a terrestrial biogeochemical model⁹, forced by simulations of transient climate change with a general circulation model¹⁰, to quantify the dynamic variations in ecosystem carbon fluxes induced by transient changes in atmospheric CO₂ and climate from 1861 to 2070. We predict that these changes increase global net ecosystem production significantly, but that this response will decline as the CO₂ fertilization effect becomes saturated and is diminished by changes in climatic factors. Thus terrestrial ecosystem carbon fluxes both respond to and strongly influence the atmospheric CO₂ increase and climate change.

Ecosystem carbon fluxes are controlled by the processes of photosynthesis, plant (autotrophic) respiration and soil (heterotrophic) respiration. The difference between plant photosynthesis and respiration is defined as net primary production (NPP), and the difference between NPP and soil respiration, defined as net ecosystem production (NEP), represents the net carbon flux from the atmosphere to ecosystems. Among many factors affecting these processes, the most obvious at the global scale are elevated CO₂ concentration and climate change, which directly and indirectly influence and interact to control the carbon fluxes from ecological and physiological processes^{1,2}. We have used the carbon exchange between vegetation, soil and the atmosphere (CEVSA) model⁹ to assess the dynamic responses of natural terrestrial ecosystems to CO₂ increases and climate change from the onset of industrialization (the 1860s) to a doubling of current radiative forcing, which will be reached in the 2060s based on IPCC scenario IS92a (ref. 11). The model simulates the processes of stomatal conductance, evapotranspiration, plant photosynthesis and respiration, nitrogen uptake, carbon allocation among plant organs, litter production, nitrogen mineralization and soil organic carbon decomposition, and uses these to calculate the carbon fluxes between vegetation, soils and the atmosphere. We do not include the effects of nitrogen deposition and modifications to ecosystem structure and distribution by natural and anthropogenic processes.

The scenarios of transient CO₂ concentration and climate change between 1861 and 2070 used in this study were produced by the Hadley Centre for Climate Prediction and Research, UK Meteorological Office with a coupled ocean–atmospheric general circulation model (GCM), which accounts for the effects of both greenhouse gases and aerosols^{10,12}. The concentrations of greenhouse gases and aerosols used by the GCM for the period 1861–1990 were derived from observed data, whereas those after 1990 were based on IPCC scenario IS92a with an increase in CO₂ by 1% yr^{−1} (ref. 11). This climate projection has been appropriately tested against observed trends and spatial patterns of climate^{10,12}.

We use the CEVSA model and the GCM scenarios of CO₂ and climate change to make simulations at a spatial resolution of 2.5° latitude × 3.75° longitude and a time step of one month. The model is first run with the GCM simulation of pre-industrial climate until an ecological equilibrium is reached, that is when the differences between annual NPP, litter production and soil respiration, and the interannual variations in soil moisture and carbon stocks were less than 0.1%. After reaching this equilibrium, the model is driven by the GCM simulation of the transient changes in CO₂ and climate from 1861 to 2070.

We isolated and combined the effects of elevated CO₂ and climate change by simulations with CO₂ increase but no climate change, climate change but no CO₂ increase, and both CO₂ increase and climate change. In addition to a global-scale analysis, the results are integrated into three latitudinal zones, northern (>40° N), temperate (40° N–20° N and >30° S) and tropical (20° N–30° S)

* Present address: Department of Environmental Sciences, University of Virginia, Charlottesville, Virginia 22903, USA.

ecosystems, to investigate the differential responses among climate-vegetation regions.

The scenarios of CO₂ increase and climate change and a comparison between the simulated and observed¹³ temperature are shown in Fig. 1. Between the 1860s and the 2060s, CO₂ increases from 288 to 640 p.p.m.v., global terrestrial temperature rises from 12.5 to 15.5 °C and annual precipitation changes little globally but increases by 9% in northern ecosystems and decreases by 4% in temperate and tropical ecosystems. The changes in climate are predicted to cause decreases in soil moisture from the 1960s (Fig. 1). In northern ecosystems increases in precipitation counteract the drying effect of warming and soil moistures change little. By contrast, in temperate and tropical ecosystems, soil moisture decreases by 10% and 8%, respectively, because of the warming and decrease in precipitation.

NPP, NEP and carbon stocks are predicted to increase substantially under CO₂ increase but no climate change (Fig. 2a). Over the period 1861–2070 NPP increases by 45% in tropical ecosystems, 36% in northern ecosystems and 20% in temperate ecosystems. Carbon stocks increase by 185 Gt in northern ecosystems, 105 Gt in temperate ecosystems and 200 Gt in tropical ecosystems. The strongest CO₂ fertilization on NPP in the tropics correlates with the largest increase in water-use efficiency. As CO₂ increases from 288 to 640 p.p.m.v., stomatal conductance and evapotranspiration in tropical ecosystems decrease by 19% and 6%, respectively,

and water-use efficiency increases by 56%. In contrast, stomatal conductance decreases by only 13% and evapotranspiration changes little in temperate and northern ecosystems, however, the water-use efficiency still increases by 31% and 42%, respectively.

Globally, the CO₂ fertilization effect is found to decrease as

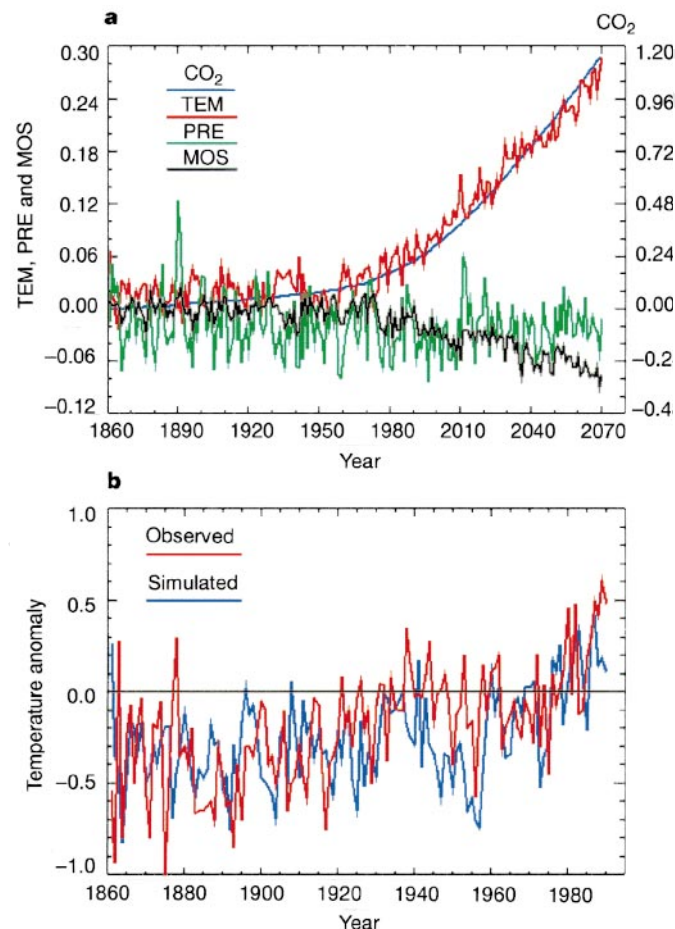


Figure 1 The scenarios. **a**, The GCM scenarios of atmospheric CO₂ and climate change¹⁰ used in this study and our simulated variation in soil moisture. All values represent the proportional changes relative to the averages in the period 1840–1880, which are 288 p.p.m.v. for atmospheric CO₂, 12.5 °C for annual mean terrestrial temperature (TEM), 805 mm for annual precipitation (PRE) and 0.51 (relative to soil water holding capacity) for soil moisture (MOS). **b**, A comparison between the GCM simulated and observed anomaly¹³ in global-mean terrestrial temperature (°C).

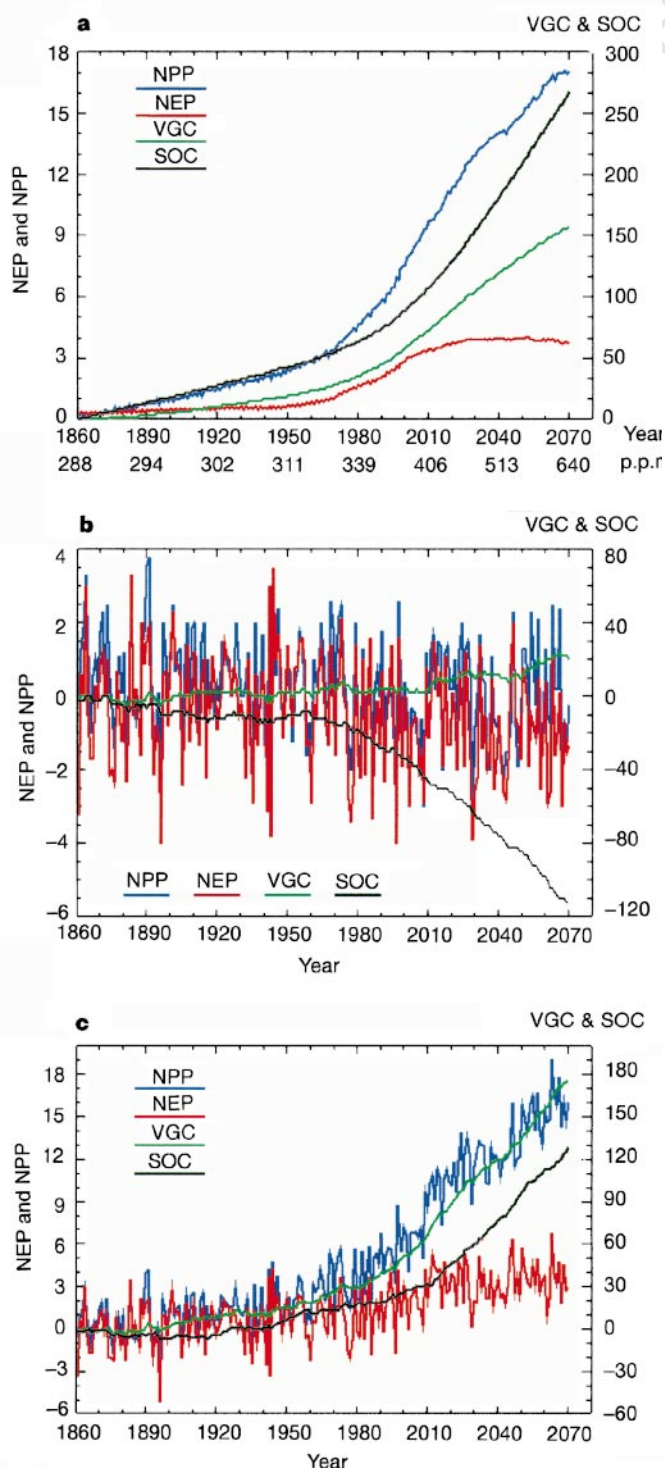


Figure 2 The variations in global net primary production (NPP, Gt C yr⁻¹), net ecosystem production (NEP, Gt C yr⁻¹), and carbon stocks in vegetation (VGC, Gt C) and soils (SOC, Gt C) resulting from atmospheric CO₂ increase and climate change. The initial values are 45 Gt C yr⁻¹ for NPP, zero for NEP, 441 Gt C for VGC, and 1238 Gt C for SOC. **a**, CO₂ increase only; **b**, climate change only; **c**, CO₂ increase and climate change.

CO₂ increases (Fig. 2a); the relative NPP enhancement decreases markedly as CO₂ exceeds 500 p.p.m.v., and NEP saturates at a CO₂ concentration of 450 p.p.m.v. The stimulation of photosynthesis by elevated CO₂ diminishes at high CO₂ concentration, but respiration increases as carbon accumulates in vegetation and soils. If CO₂ increases continually, NPP is predicted to saturate at a CO₂ concentration of about 1,200 p.p.m.v., and NEP declines as CO₂ exceeds 600 p.p.m.v. (Fig. 3). If CO₂ is stabilized, NEP falls rapidly and eventually reaches zero (Fig. 3), because photosynthesis will be balanced by increased plant and soil respiration.

Under climate change with no CO₂ increase, we predict wide interannual fluctuations in the carbon fluxes (Fig. 2b) and differential responses among various ecosystems. In northern ecosystems, NPP is enhanced by 13% and vegetation carbon stocks are increased by 28 Gt C in the period 1861–2070, however, soil carbon stocks are reduced by 55 Gt C at the same time because soil respiration is stimulated by warming. NEP is predicted to decrease to –0.2 Gt C yr^{–1} in the 1980s, and to –0.5 Gt C yr^{–1} in the 2060s. In temperate ecosystems NPP and NEP are slightly reduced with a total loss of 8 Gt C between 1861 and 2070. In tropical ecosystems, NPP is reduced by 6% and NEP is decreased to about –0.4 Gt C yr^{–1} from the 1970s, leading to a total loss of 52 Gt C in the period 1861–2070.

Climate change has little effect on global NPP, the increase in the north offsets the decrease in the tropics. However, global NEP is decreased to –0.4 Gt C yr^{–1} in the period 1951–2000, and to –0.8 Gt C yr^{–1} in the period 2001–2070, consistent with a predicted global reduction in soil moisture (Fig. 1a). Although vegetation carbon stocks are increased slightly, soil carbon stocks are reduced markedly (Fig. 2b). The total carbon stocks decrease by 87 Gt from 1861 to 2070 because of stronger increases in soil respiration than NPP in the north and decreases in NPP in the tropics. This result is supported by estimates that the effect of warming on respiration dominates that on plant growth and reduces soil carbon storage^{14,15}, although it is contrary to some suggestions that climate change has caused a substantial carbon sink^{16,17}. Warming may enhance NPP in the north as suggested by studies^{17,18} based on satellite data of NDVI (the normalized difference vegetation index), but it does not necessarily increase the carbon sink because warming also stimulates soil respiration.

When both CO₂ concentration and climate change, NPP, NEP and carbon stocks all increase (Fig. 2c). From 1861 to 2070 global NPP increases by 36%. Global NEP varies between –6.5 and 5.3 Gt C yr^{–1}, responding positively to changes in CO₂ and precipitation (PRE) but negatively to changes in temperature (TEM) (NEP = 0.036[CO₂] – 2.431TEM + 0.029PRE – 2.868, $r^2 = 0.64$). NEP increases significantly from the 1950s along with the rapid increase

in CO₂, but it steadies after 2020 (Fig. 2c) as the CO₂ fertilization effect saturates (Fig. 2a) and is diminished by climate change (Fig. 2b). The decadal-mean NEP becomes positive from the 1910s and increases to 1.1 Gt C yr^{–1} in the period 1950–1990, and to 3.2 Gt C yr^{–1} in the period 2030–2070. This causes an increase of 309 Gt C in soils and vegetation from 1861 to 2070, in contrast with an increase of 490 Gt C under CO₂ increase and a loss of 87 Gt C under climate change alone.

Increases in CO₂ and temperature interact positively to enhance NPP in the north, whereas elevated CO₂ reverses the decreases in NPP caused by climate change in the tropics. Marked increases in NEP therefore occur in both the tropics and the north under combined changes in CO₂ and climate (Fig. 4a). The NEPs in the north and tropics are of a similar magnitude under the contemporary climate, but NEPs in the north will become much higher than in the tropics in the future (Fig. 4a) because the strongest NPP increase is predicted to occur at northern high latitudes. Although various evidence supports a substantial carbon sink in the north, data confirming a tropical carbon sink are limited^{3,4}. Direct measurements and modelling of CO₂ fluxes^{6,19} show that tropical ecosystems are taking up carbon; however, the carbon sink may be concealed by the carbon release from deforestation⁴.

Climate change and elevated CO₂ also modify seasonal ecosystem carbon fluxes. NEP increases between May and September but decreases in other months, its seasonal amplitude is therefore greatly amplified (Fig. 4b). This variation is mainly due to the changes occurring in the north: the increase in soil respiration is higher than in NPP in winter and spring but is lower in summer. Strong increases in soil respiration by warming in winter and spring have been observed^{20,21}. In summer, the advantage of increased

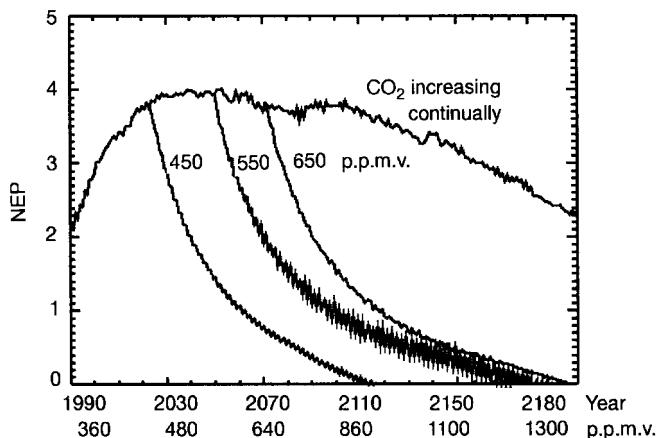


Figure 3 Variations in global net ecosystem production (NEP, Gt C yr^{–1}) responding to stabilization of atmospheric CO₂ at 450, 550 and 650 p.p.m.v., or continual increase. This estimate is made with changes in atmospheric CO₂ only.

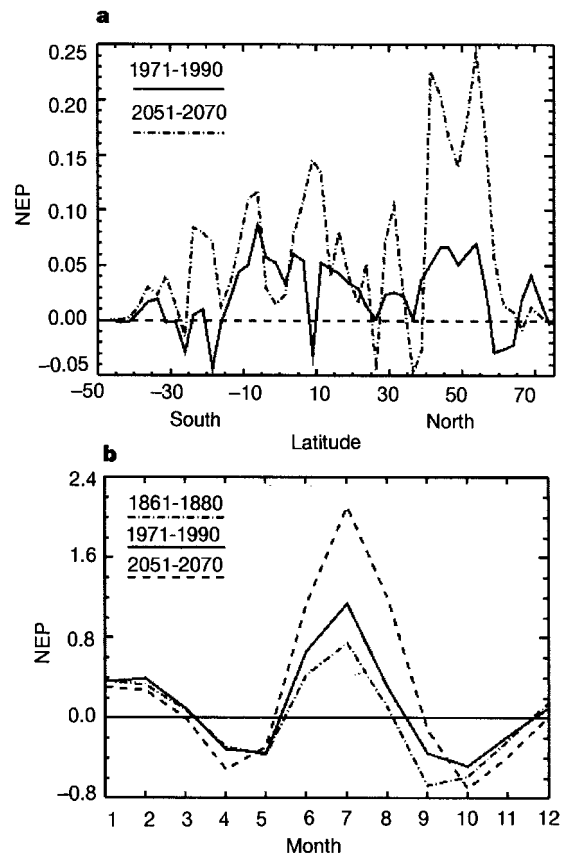


Figure 4 Spatial and seasonal variations in net ecosystem production (NEP) in response to atmospheric CO₂ increase and climate change. **a**, Latitudinal distributions in NEP (Gt C yr^{–1}) in the periods 1971–1990 and 2051–2070. All values represent that for the land in 2.5° latitudinal bands. **b**, Seasonal change in global NEP (Gt C month^{–1}) in the periods 1861–1880, 1971–1990 and 2051–2070.

temperature and CO₂ to NPP can be fully utilized because precipitation is plentiful and leaf canopies are fully expanded.

A consequence of CO₂ increase and climate change is that global terrestrial ecosystems are estimated to have sequestered 58 Gt C in the period 1861–1990, accounting for about two-thirds of the unattributed terrestrial biospheric sink from deconvolutions²². Forest regrowth^{23,24} and nitrogen deposition^{7,25} also contribute to the terrestrial carbon sink, although estimates of these sinks are particularly uncertain. Forests accumulate carbon in young and middle age, but the rates decline to zero as stands mature. A full inventory of forest demographic changes is essential to estimate the carbon sinks caused by forest regrowth. Although disturbance by human (such as harvest and management) and natural factors (such as fire and wind) initially stimulate forest growth^{23,24}, decreased nutrient availability and increased stomatal limitation cause a decline in NPP as stands age²⁶. Nitrogen is the primary nutrient that limits both plant production and soil decomposition⁹ and often constrains CO₂ fertilization²⁷. Plants may acclimatize to nitrogen supply by adjusting nitrogen allocation among organs and biomass distribution between under- and aboveground^{7,27}. These mechanisms are central to estimates of the effect of nitrogen deposition and the interaction with elevated CO₂, but little is known at present. It is important to recognize also that changes in CO₂ and climate may cause transient changes in vegetation composition and distribution²⁸, but this effect is not included in this study. The sensitivity of vegetation to CO₂ concentration and climate varies with plant species and stand age^{26,28}, and rapid climate change can cause the die back of some vegetation types and consequent carbon loss before new types are established⁸. Models simulating ecosystem structure dynamics are being developed to assess the effects of demographic and compositional changes of forests^{29,30}.

We predict that the net effect of CO₂ increase and climate change is to increase global NEP substantially, but that the response of natural terrestrial ecosystems depends largely upon changes in individual factors and varies widely among climate-vegetation zones. This information is critical for calculations of CO₂ stabilization and new developments in GCMs, in which the feedback of ecosystem carbon cycling on CO₂ have not yet been taken into account. The dynamic simulation of ecosystem carbon fluxes in this study, along with progress in understanding the effect of nitrogen deposition and transient changes in ecosystem structure, will lead to a full coupling between ecosystem carbon processes and climatic systems that is essential for more realistic predictions of both ecosystem responses and climate change. □

Received 5 June 1997; accepted 16 March 1998.

1. Amthor, J. S. Terrestrial higher-plant response to increasing atmospheric [CO₂] in relation to the global carbon cycle. *Global Change Biol.* **1**, 243–247 (1995).
2. Houghton, R. A. & Woodwell, G. M. Global climatic change. *Sci. Am.* **260**, 36–47 (1989).
3. Ciais, P., Tan, P. P., Troler, M., White, J. W. C. & Francy, R. J. A large northern hemisphere terrestrial CO₂ sink indicated by ¹³C/¹²C of atmospheric CO₂. *Science* **269**, 1098–1102 (1995).
4. Keeling, R. F., Piper, S. C. & Heimann, M. Global and hemispheric CO₂ sinks deduced from changes in atmospheric O₂ concentration. *Nature* **381**, 218–221 (1996).
5. Wofsy, S. C., Munger, J. E., Bakwin, P. S., Daube, B. C. & Moore, T. R. Net CO₂ uptake by northern woodlands. *Science* **260**, 1314–1317 (1993).
6. Grace, J. et al. Carbon dioxide uptake by undisturbed tropical forests, 1992 and 1993. *Science* **270**, 778–780 (1995).
7. Melillo, J. M. in *Global Change and Terrestrial Ecosystems* (eds Walker, B. & Steffen, W.) 431–450 (Cambridge Univ. Press, 1996).
8. Smith, T. M. & Shugart, H. H. The transient response of terrestrial carbon storage to a perturbed climate. *Nature* **361**, 523–526 (1993).
9. Cao, M. K. & Woodward, F. I. Net primary and ecosystem productions and carbon stocks of terrestrial ecosystems and their response to climate change. *Global Change Biol.* **4**, 185–198 (1998).
10. Mitchell, J. F. B., Johns, T. C., Gregory, J. M. & Tett, S. F. B. Climate response to increasing levels of greenhouse gases and sulphate aerosols. *Nature* **376**, 501–504 (1995).
11. Houghton, J. T., Callander, B. A. & Varney, S. K. (eds) *Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment* (Cambridge Univ. Press, 1992).
12. Johns, T. C. et al. The second Hadley Centre coupled ocean-atmosphere GCM: model description, spinup and validation. *Clim. Dyn.* **13**, 103–134 (1997).
13. Jones, P. D. Hemispheric surface air temperature variations: a reanalysis and an update to 1993. *J. Clim.* **7**, 1794–1802 (1994).
14. Oechel, W. C. et al. Recent change of Arctic tundra ecosystems from a net carbon dioxide sink to a source. *Nature* **361**, 520–523 (1993).
15. Townsend, A. R., Vitousek, P. M. & Holland, E. A. Tropical soils could dominate the short-term carbon cycle feedback to increased global temperature. *Clim. Change* **22**, 293–303 (1992).

16. Dai, A. & Fung, I. Y. Can climate variability contribute to the “missing” CO₂ sink? *Glob. Biogeochem. Cycles* **7**, 599–609 (1993).
17. Braswell, B. H., Schimel, D. S., Linder, E. & Moore, B. The response of global terrestrial ecosystems to interannual temperature variability. *Science* **278**, 870–873 (1997).
18. Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G. & Nemani, R. R. Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature* **386**, 698–702 (1997).
19. Lloyd, J. et al. A simple calibrated model of Amazonian rainforest productivity based on leaf biochemical properties. *Plant Cell Environ.* **18**, 1129–1145 (1995).
20. Oechel, W. C., Vourlitis, G. & Hastings, S. J. Cold season CO₂ emission from Arctic soils. *Glob. Biogeochem. Cycles* **11**, 151–162 (1997).
21. Zimov, S. A. et al. Siberian CO₂ efflux in winter as a CO₂ source and cause of seasonality in atmospheric CO₂. *Clim. Change* **33**, 111–120 (1996).
22. Houghton, R. A. in *The Global Carbon Cycle* (ed. Heimann, M.) 139–157 (Springer, New York, 1993).
23. Dixon, R. K. et al. Carbon pools and flux of global forest ecosystems. *Science* **263**, 185–190 (1994).
24. Melillo, J. M., Fruce, P. A., Houghton, R. A., Moore, B. & Skole, D. L. Land-use changes in the Soviet Union between 1850 and 1980: causes of a net release of CO₂ to the atmosphere. *Tellus B* **40**, 116–128 (1988).
25. Holland, E. A. et al. Variations in the predicted spatial distribution of atmospheric nitrogen deposition and their impact on carbon uptake by terrestrial ecosystems. *J. Geophys. Res.* **102**, 15849–15866 (1997).
26. Gower, S. T., McMurtrie, R. E. & Durty, D. Aboveground net primary production decline with stand age: potential causes. *Trends Ecol. Evol.* **11**, 378–383 (1996).
27. McGuire, A. D., Melillo, J. M. & Joyce, L. A. The role of nitrogen in the response of forest net primary production to elevated atmospheric carbon dioxide. *Annu. Rev. Ecol. Syst.* **26**, 473–503 (1995).
28. Schulze, E.-D. Flux control at the ecosystem level. *Trends Ecol. Evol.* **10**, 40–43 (1994).
29. Foley, J. A. et al. An integrated biosphere model of land-surface processes, terrestrial carbon balance, and vegetation dynamics. *Glob. Biogeochem. Cycles* **10**, 603–628 (1996).
30. Pitelka, L. F. et al. Plant migration and climate change. *Am. Sci.* **85**, 464–473 (1997).

Acknowledgements. This work is being supported by the Natural Environment Research Council, UK. We thank R. Betts, the Hadley Centre for Climate Prediction and Research, UK for supplying the climatic and CO₂ data, P. L. Mitchell and D. J. Beerling for their critical comments.

Correspondence and request for materials should be addressed to M.C. (e-mail: mc5m@virginia.edu).

Mineralogy and dynamics of a pyrolite lower mantle

S. E. Kesson, J. D. Fitz Gerald & J. M. Shelley

Research School of Earth Sciences, Australian National University, Canberra ACT 0200, Australia

There is a growing consensus that the Earth's lower mantle possesses a bulk composition broadly similar to that of the upper mantle (known as pyrolite)^{1–3}. But little is known about lower-mantle mineralogy and phase chemistry^{4,5}, especially at depth. Here we report diamond-anvil cell experiments at pressures of 70 and 135 GPa (equivalent to depths within the Earth of about 1,500 and 2,900 km, respectively) which show that pyrolite would consist solely of magnesian-silicate perovskite (MgPv), calcium-silicate perovskite (CaPv) and magnesiowüstite (Mw). Contrary to recent speculation^{6,7}, no additional phases or disproportionations were encountered and MgPv was found to be present at both pressures. Moreover, we estimate that, at ultra-high pressures where thermal expansivities are low, buoyancy forces inherent in subducted slabs because of their lithology will be of similar magnitude to those required for thermally driven upwelling. So slabs would need to be about 850 °C cooler than their surroundings if they are to sink to the base of the mantle. Furthermore, initiation of plume-like upwellings from the core-mantle boundary, long attributed to superheating, may be triggered by lithologically induced buoyancy well before thermal equilibration is attained. We estimate that ascent would commence within ~0.5 Gyr of the slab reaching the core-mantle boundary, in which case the lowermost mantle should not be interpreted as a long-term repository for ancient slabs.

The diamond-anvil cell methodology and its inherent limitations are described in Fig. 1 legend. Our results constrain both the mineralogy and phase chemistry that would be adopted at depth by a pyrolite lower mantle, and cannot substantiate a report⁷ that MgSiO₃ perovskite disproportionates to its component oxides at pressures near 70 GPa. Moreover, existence of the three-phase assemblage of MgPv + Mw + CaPv at ~135 GPa means that no

Table 1 Representative phase chemistry of pyrolite at 70 and 135 GPa

	MgPv 70 GPa	Mw 70 GPa	CaPv 70 GPa	MgPv 135 GPa	Mw 135 GPa	CaPv 135 GPa
SiO ₂	56.9	n.d.	54.0	55.4	n.d.	51.2
TiO ₂	0.2	n.d.	n.d.	0.5	n.d.	tr.
Al ₂ O ₃	5.6	1.2	0.7	5.0	0.8	1.0
Cr ₂ O ₃	0.5	1.1	n.d.	0.5	0.8	n.d.
MnO	n.d.	tr.	n.d.	n.d.	tr.	tr.
FeO	4.7	23.0	1.9	4.5	24.0	2.2
MgO	31.1	71.5	3.7	32.9	71.9	6.1
NiO	n.d.	0.5	n.d.	n.d.	0.4	n.d.
CaO	0.6	n.d.	37.7	1.2	n.d.	36.6
Na ₂ O	tr.	2.1	n.d.	n.d.	1.6	0.9
K ₂ O	n.d.	n.d.	1.6	n.d.	n.d.	1.8
Mg#	92.2	84.7		92.8	84.2	

Analytical uncertainties are about 1% relative for all elements. A portion of the NiO inventory is unaccounted for. It has probably been reduced to Fe–Ni alloy to stabilize minor amounts of ferric iron as required by the major phases. Abbreviations: n.d., below detection; tr., qualitatively discernible at levels near the detection limit. Mg# = 100 MgO/(MgO + FeO) molar.

new aluminous phases are stabilized at depth⁶. Furthermore the seismically anomalous D'' layer overlaying the core should not be ascribed to a fortuitous disproportionation.

After thermal quenching and relaxation of pressure, MgPv is found to be orthorhombic (*Pbmn*), Mw has the rocksalt structure and CaPv is, as always, amorphous. Subsolidus phase chemistry for pyrolite is remarkably uniform and insensitive to temperature or pressure, as seen from representative analyses in Table 1. As reported elsewhere⁸, NiO and Cr₂O₃ partition into Mw. The unexpected compatibility of Na₂O in Mw (ref. 5) is confirmed, and an Na_{0.5}M_{0.5} component in solid solution is implicated (M = Cr³⁺ + Al³⁺ ± Fe³⁺). Most of the Al₂O₃ inventory, however, is accommodated in

MgPv (~5 wt%). Ti⁴⁺ partitions into MgPv rather than CaPv, although CaPv is the host for K₂O, and also contains traces of MnO and Na₂O at a pressure of 135 GPa. Solid solution of CaPv in MgPv is minimal, whereas limited solid solution (~20 mol%) of MgPv in CaPv is observed at the highest pressures.

A literature survey confirms that partitioning of ferrous iron (Fe²⁺) between MgPv and Mw is poorly constrained (see, for example, ref. 9). Figure 2 shows that *K_D* is relatively insensitive to pressure and temperature, averaging about 0.50 ± 0.05 at 70 GPa and 0.42 ± 0.05 at 135 GPa. These *K_D* values overlap their counterparts in an Al₂O₃-free system (Fo₉₀, shaded band). In principle, the aluminous character of MgPv should alter its partition behaviour by facilitating the substitution of Mg²⁺ by Fe²⁺ (refs 5, 10). However, the limited alumina inventory of lower-mantle MgPv is ineffective in this context (Fig. 2). We therefore use just one *K_D* value (0.45) for both fertile and depleted peridotite in subsequent calculations.

Densities of a pyrolite lower mantle, calculated as described in Table 2, agree closely with the preliminary reference Earth model PREM¹¹. Unfortunately, we cannot discriminate between competing compositional models for the lower mantle using density alone because the uncertainties accompanying extrapolation of thermoelastic parameters to lower mantle pressure and temperature conditions are prohibitive in this context. We can, however, model density differences between subducted slabs and their lower mantle surroundings with more confidence because both incorporate peridotitic lithologies. Errors attributable to MgPv, Mw and CaPv are systematic and are thereby partly eliminated (Tables 2 and 3). However, the validity of our conclusions ultimately relies on the fact that thermal expansivities at ultra-high pressures are very small indeed¹². Buoyancy forces accordingly require large lateral temperature variations. Consequently, buoyancy that is inherent in the very nature of a slab by virtue of its lithology—basalt underlain by depleted peridotite—becomes increasingly important as pressure

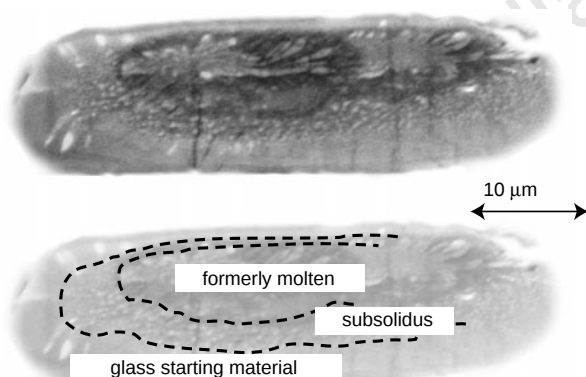


Figure 1 Polished cross-section of a DAC charge recovered from a pressure of 70 GPa. Experimental starting materials were synthetic homogenous ultramafic glass fragments of Ringwood's pyrolite composition²⁸ prepared in small quantities by high-speed quenching techniques. (The same oxide-mix precursor had been used⁶ to characterize pyrolite at 28 GPa.) Glass fragments were loaded into steel gaskets and compressed at 70 (±5) and 135 (±10) GPa in a Bohler-design diamond-anvil high-pressure cell. Pressure was measured by ruby-fluorescence; experimental procedures have been described elsewhere^{8,16,29}. Heating was induced by the infrared beam of a stabilized YAG laser. The beam, incident onto the top surface of the sample, was defocused to minimize lateral temperature gradients, and left for 0.5 h in the same position (but temperature gradients and deviatoric stresses are nevertheless far more severe than those inside the Earth). An ovoid transformed volume ~50 μm in diameter and 12 μm thick was formed (top). Backscattered electron imaging shows that the central portion of the transformed ovoid has melted, whereas peripheral regions remained subsolidus. Ion-beam erosion was used to expose transformed material for transmission electron microscopy (TEM). Sequential ion-erosion parallel to the long dimension allowed us to characterize the temperature history from the beginning of nucleation through to the indecipherably complex, formerly molten region. Thus a zone of subsolidus transformation products has experienced temperatures appropriate for the geotherm even though we could not control or measure temperature. Crystallographic and chemical data were obtained by TEM^{8,16,29} on quenched, recovered specimens.

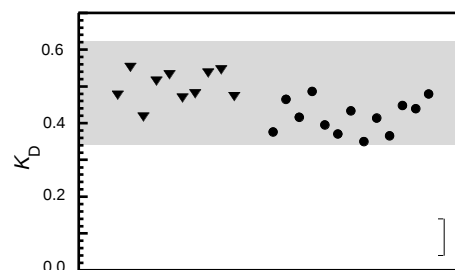


Figure 2 Partitioning of iron between pyrolite MgPv and Mw pairs in contact (*K_D*). *K_D* = molar (FeO_{MgPv}/MgO_{MgPv})/(FeO_{Mw}/MgO_{Mw}). Coefficients averaging 0.50 ± 0.05 at 70 GPa (triangles), and 0.42 ± 0.05 at 135 GPa (circles), overlap our unpublished data for their counterparts in an Al₂O₃-free system (Fo₉₀) at ~100 GPa and similar temperatures (shaded band). Mean grain size, and therefore temperature, increase from left to right across each subset.

Table 2 Thermoelastic parameters, mineral densities and proportions

Phase or lithology	V_0 (cm ³)	K_0 (GPa)	K'_0	θ_0	γ_0	q	ρ at 1,100 km (g cm ⁻³)	ρ at CMB (g cm ⁻³)	wt% in Pyrolite	wt% in depleted peridotite	wt% in MORB
Mw (Mg# 87.0)	11.397	164.2	4.13	673	1.41	1.3	4.53	5.56		24	
Mw (Mg# 83.5)	11.412	164.4	4.13	673	1.41	1.3	4.62	5.68	22		
MgPv (2% Al ₂ O ₃ ; Mg# 93.7)	24.497	262.4	3.8	1,000	1.33	1.4	4.62	5.49		71	
MgPv (5% Al ₂ O ₃ ; Mg# 92.7)	24.526	262.4	3.8	1,000	1.33	1.4	4.63	5.50	69		
MgPv (20% Al ₂ O ₃ ; Mg# 71.0)	25.139	260	3.8	1,000	1.33	1.4	4.71	5.60			36
CaPv	27.453	232.0	4.8	1,000	1.33	1.0	4.73	5.56	9	5	31
Stishovite	14.080	305	5.3	1,149	1.35	1.0	4.61	5.28			14
Ca-ferrite (Mg,Fe,Na)AlSiO ₄ ss	35.576	241	4.0	900	1.3	1.0	4.60	5.48			19
PREM ¹¹							4.64	5.56			
Pyrolite							4.64	5.55			
MORB							4.68	5.52			
Depleted peridotite							4.61	5.50			
Thermally equilibrated slab							4.62	5.51			

Mineral densities are from refs 2, 20–25, and proportions are from ref. 16 and our results. Densities were calculated following the methodology of ref. 25. The slab is a model comprised of two representative lithologies. Depleted peridotite complementary to mid-ocean-ridge basalt (MORB) was obtained by extracting 20 wt% of a suitably primitive MORB partial melt²⁶ from fertile upper mantle²⁷. The model slab comprises MORB plus depleted peridotite (20:80 wt% respectively). Abbreviations: V_0 , molar volume; K_0 , isothermal bulk modulus; K'_0 , first pressure derivative of K_0 ; θ_0 , Debye temperature reference value at zero pressure and room temperature; γ_0 , Grüneisen parameter at zero pressure; $q = (\delta \ln \gamma / \delta \ln V)_T$; ρ , density. Temperature and pressure at 1,100 km are 2,050 K and 43.2 GPa, and at CMB are 2,650 K and 135 GPa; MORB densities incorporate new data²⁸ and so supersede our earlier estimates¹⁶; PREM, preliminary reference Earth model¹¹.

Table 3 Thermal retardation and density deficits

Lithology	1,100 km		CMB	
	Density difference relative to pyrolite (%)	Thermal retardation (°C)	Density difference relative to pyrolite (%)	Thermal retardation (°C)
Thermally equilibrated former basalt (MORB)	+0.9 ± 0.1		−0.5 ± 0.1	
Thermally equilibrated depleted peridotite	−0.8 ± 0.1		−0.7 ± 0.1	
Thermally equilibrated slab	−0.4 ± 0.1	0	−0.6 ± 0.1	0
Neutrally buoyant slab	0	200–300	0	550–750
Thermally retarded, sinking slab	+0.2 ± 0.1	300–400	+0.2 ± 0.1	750–950

Overall uncertainties of ±0.1% were evaluated using MgPv in the first instance. K' was raised to 4.0, K_0 changed by 4% relative and q allowed to range from 1 to 2. MgPv density, calculated²⁸ at CMB, remains within ±0.5% relative. Errors of this same magnitude are then ascribed to all mineral densities at CMB, and density differences between slab and mantle calculated accordingly. (Note that the sense of error has to be the same for all MgPv simultaneously; the same is true for Mw.)

increases.

Tomographic imaging shows that many slabs, despite variable lateral deflection within the transition zone, eventually penetrate the lower mantle and reach depths of at least 1,500 km. Seismic velocity anomalies in these slabs imply that they are thermally retarded (substantially cooler than their immediate surroundings). Thereafter, limitations in the database mean that the global picture is obscure¹³, although at least two such slabs, the Japanese and Farallon plates, are now known to reach the core–mantle boundary (CMB)^{13,14}.

The basaltic crust of slabs finally transforms to perovskite at depths of about 1,100 km (refs 15, 16). We calculate that a state of neutral buoyancy would be attained by a slab that is about 250 ± 50 °C cooler than its surroundings at 1,100 km, and approximately 350 °C cooler at 1,500 km. Neutrally buoyant slabs would not continue settling but would instead recirculate within the lower mantle's convective system. Such entrainment might account for the disappearance of many slabs from tomographic view in the lower reaches of the mantle and for lateral heterogeneity in seismic properties.

The inherent buoyancy of slabs is at a maximum at depths equivalent to the CMB. At such depths, neutral buoyancy is attained when slabs are about 650 ± 100 °C cooler than surrounding pyrolite. So what are the requirements for a slab to settle to the CMB? If negative buoyancy of 0.2%, for example, is necessary for sinking, then at 1,100 km that is equivalent to an average thermal retardation of only 350 ± 50 °C. But if a slab is to reach the CMB it should be about 850 ± 100 °C cooler than its surroundings. Perhaps only fast-moving plates meet that criterion.

The proportions of basalt and peridotite that eventually reach the lower mantle might be altered in transit by dynamic processes. The majority consensus, judging from the literature, is that the profound buoyancy of former basalt at depths of between about 600 and 1,100 km might result in its delamination and entrapment within the transition zone¹⁷. This would in turn enhance the proportion of lighter peridotite lithologies in any slab that subsequently penetrated

into the lower mantle.

The inherent buoyancy of slabs at CMB depths means that the D'' layer cannot form a long-term repository for former basalt and/or foundered slabs¹⁸. Rather, the opposite is the case because reheating will be most effective at the core–mantle interface, and both peridotite and, probably, basalt become buoyant well before they reach thermal equilibrium (Table 3). Moreover, simple heat-transfer calculations indicate that, from the perspective of isotopic evolution, residency at CMB depths would be brief, no more than about 0.5 Gyr. These considerations challenge both current perceptions of the dynamics of plume ascent and the view that the D'' layer is a repository for ancient subducted lithosphere.

The dynamic implications of reheating at the base of the lower mantle deserve further emphasis. Thermal expansivities at ultra-high pressures¹² are so small (about 10^{−5} K^{−1}) that lithologically induced density deficits are of the same magnitude as those required for thermally driven convection. For example, the density deficit between a thermally equilibrated slab and its CMB surroundings is equivalent to superheating pyrolite by ~600 °C, more than enough to trigger upwelling. (It has been estimated¹⁹ that superheating of ~400 °C would confer sufficient buoyancy for plume ascent in a compositionally uniform system.) Initiation of plumes may therefore be driven by petrological factors, rather than excess temperature, with thermal driving forces being required only for the continued feeding of plume tails. □

Received 12 February 1997; accepted 18 March 1998.

- Anderson, O. L. Finding the isentropic density of perovskite: implications for iron concentration in the lower mantle. *Geophys. Res. Lett.* **24**, 213–216 (1997).
- Jackson, I. & Rigden, S. M. in *The Earth's Mantle: Composition, Structure and Evolution* (ed. Jackson, I.) 405–460 (Cambridge Univ. Press, 1998).
- Stacey, F. D. Thermoelasticity of (Mg,Fe)SiO₃ perovskite and a comparison with the lower mantle. *Phys. Earth Planet. Inter.* **98**, 65–77 (1996).
- O'Neill, B. & Jeanloz, R. Experimental petrology of the lower mantle: A natural peridotite taken to 54 GPa. *Geophys. Res. Lett.* **77**, 1477–1480 (1990).
- Irfune, T. Absence of an aluminous phase in the upper part of the Earth's lower mantle. *Nature* **370**, 131–133 (1994).
- Ahmed-Zaid, I. & Madon, M. Electron microscopy of high-pressure phases synthesized from natural

- garnets, kyanite and anorthite in a diamond anvil cell. *Earth Planet. Sci. Lett.* **129**, 233–247 (1995).
7. Saxena, S. K. *et al.* Stability of perovskite (MgSiO₃) in the Earth's mantle. *Science* **274**, 1357–1359 (1996).
 8. Kesson, S. E. & Fitz Gerald, J. D. Partitioning of MgO, FeO, NiO, MnO and Cr₂O₃ between magnesian silicate perovskite and magnesio-wüstite: implications for the origin of inclusions in diamond and the composition of the lower mantle. *Earth Planet. Sci. Lett.* **111**, 229–240 (1991).
 9. Mao, H. K., Shen, G. & Hemley, R. J. Multivariable dependence of Fe–Mg partitioning in the lower mantle. *Science* **278**, 2098–2100 (1997).
 10. Wood, B. J. & Rubie, D. C. The effect of alumina on phase transformations at the 660 km discontinuity from Fe–Mg partitioning experiments. *Science* **273**, 1522–1524 (1996).
 11. Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
 12. Chopelas, A. & Boehler, R. Thermal expansivity in the lower mantle. *Geophys. Res. Lett.* **19**, 1983–1986 (1992).
 13. van der Hilst, R., Widiyantoro, S. & Engdahl, E. R. Evidence for deep mantle circulation from global tomography. *Nature* **386**, 578–584 (1997).
 14. Grand, S. & Helmberger, D. V. Upper mantle shear structure of North America. *Geophys. J. R. Astron. Soc.* **76**, 399–438 (1984).
 15. Faust, J. & Knittle, E. The stability and equation of state of majoritic garnet synthesised from natural basalt at mantle conditions. *Geophys. Res. Lett.* **23**, 3377–3380 (1996).
 16. Kesson, S. E., Fitz Gerald, J. D. & Shelley, J. M. Mineral chemistry and density of subducted basaltic crust at lower mantle pressures. *Nature* **372**, 767–769 (1994).
 17. Irifune, T. & Ringwood, A. E. Phase transformations in subducted oceanic crust and buoyancy relationships at depths of 600–800 km in the mantle. *Earth Planet. Sci. Lett.* **117**, 101–110 (1993).
 18. Christensen, U. R. & Hofmann, A. W. Segregation of subducted oceanic crust in the convecting mantle. *J. Geophys. Res.* **89**, 4389–4402 (1994).
 19. Griffiths, R. & Campbell, I. Stirring and structure in mantle plumes. *Earth Planet. Sci. Lett.* **99**, 66–78 (1990).
 20. Wang, Y., Weidner, D. J., Liebermann, R. C. & Zhao, Y. P-V-T equation of state of (Mg,Fe)SiO₃ perovskite: constraints on composition of the lower mantle. *Phys. Earth Planet. Inter.* **83**, 13–40 (1994).
 21. Wang, Y., Weidner, D. J. & Guyot, F. Thermal equation of state of CaSiO₃ perovskite. *J. Geophys. Res.* **101**, 661–672 (1996).
 22. Yutani, M., Yagi, T., Yusa, H. & Irifune, T. Compressibility of calcium ferrite-type MgAl₂O₄. *Phys. Chem. Minerals* **24**, 340–344 (1997).
 23. Anderson, O. L., Masuda, K. & Isaak, D. G. Limits on the value of δ_T and γ for MgSiO₃ perovskite. *Phys. Earth Planet. Inter.* **98**, 31–46 (1996).
 24. Li, B., Rigden, S. M. & Liebermann, R. C. Elasticity of stishovite at high pressure. *Phys. Earth Planet. Inter.* **96**, 113–128 (1996).
 25. Ita, J. & Stixrude, L. Petrology, elasticity and composition of the mantle transition zone. *J. Geophys. Res.* **97**, 6849–6866 (1992).
 26. Niu, Y. & Batiza, R. An empirical method for calculating melt compositions produced below mid-ocean ridges; application for axis and off-axis (seamounts) melting. *J. Geophys. Res.* **96**, 21753–21777 (1991).
 27. O'Neill, H. S. C. & Palme, H. in *The Earth's Mantle: Composition, Structure and Evolution* (ed. Jackson, I.) 3–128 (Cambridge Univ. Press, 1998).
 28. Ringwood, A. E. *Origin of the Earth and Moon* (Springer, New York, 1979).
 29. Kesson, S. E., Fitz Gerald, J. D. & Shelley, J. M. G. Phase relations, structure and crystal chemistry of some aluminous silicate perovskites. *Earth Planet. Sci. Lett.* **134**, 187–201 (1995).

Acknowledgements. We thank I. Jackson, H. St C. O'Neill, G. F. Davies and A. Leitch for comments.

Correspondence and requests for materials should be addressed to S.K. (Sue.Kesson@anu.edu.au).

Ichthyosaurian relationships illuminated by new primitive skeletons from Japan

Ryosuke Motani*, Nachio Minoura† & Tatsuro Ando†

* University of California Museum of Paleontology, 1101 VLSB, Berkeley, California 94720-4780, USA

† Department of Earth and Planetary Science, Hokkaido University, Sapporo 060, Japan

The Ichthyosauria is a group of reptiles with fish-shaped bodies from the Mesozoic (65–250 million years ago)^{1,2}. Their secondary adaptations to aquatic life have obscured their ancestral features^{3,4}, and basal ichthyosaurs, which would be expected to retain these ancestral features (plesiomorphies), are poorly represented in the fossil record¹. As a result, their relationships to other amniotes have been controversial for over 180 years^{5,6}. New specimens of *Utatsusaurus hataii* from the Lower Triassic (240 Myr ago) of Japan are the first basal ichthyosaurs to show detailed features for almost the entire skeleton, including previously unknown parts of the skull and pelvic girdle. Computer-assisted retrodeformation of fossil images⁷ shows that *Utatsusaurus* retained features of terrestrial amniotes in both the skull and the postcranial skeleton, such as the connection between the vertebral column and the pelvic girdle. Phylogenetic analyses indicate that ichthyosaurs belong in the Diapsida, but

that, unlike the sauropterygians^{8,9}, they are not included with the Sauria (the crown group containing lizards, crocodiles, birds and *Sphenodon*). Recent studies have reported that the addition of ichthyosaurs to the amniote data altered the relationships among basal saurians^{10,11}, but no major clades were affected by the inclusion of ichthyosaurs in our analyses.

Our study is based on nearly complete skeletons collected independently from the Lower Triassic (Spathian) of Ogatsu, Miyagi, Japan in 1982, by Hokkaido University (UHR 30691) and the National Science Museum (NSM-VP-20028). The locality is about 20 km south of the type locality of *Utatsusaurus*, where the same Osawa Formation is exposed¹². These skeletons experienced tectonic deformation which so altered their shape that they were initially considered to represent a new taxon¹³. However, two-dimensional retrodeformation of their images (see Methods) revealed close morphological similarities to the holotype of *Utatsusaurus hataii* (Fig. 1). An estimate of the length of *Utatsusaurus*, based on partial skeletons, was about 1.5 m (ref. 14). However, the new specimens show that *Utatsusaurus* was nearly 3 m long, with a slender body (Fig. 2e), as in the contemporaneous *Chensaurus*², which was one-third its size. There are approximately 40 presacral vertebrae which are cylindrical, suggesting that *Utatsusaurus* probably swam with an eel-like motion, as did *Chensaurus*.

Utatsusaurus is phylogenetically basal among ichthyosaurs¹⁵, and the specimens described here show features that are transitional between ancestral terrestrial amniotes and the more derived ichthyosaurs. For example, the species retained two well defined sacral ribs with marked distal expansions (Fig. 2c) that are unknown in any other ichthyosaurs. The sacral ribs probably articulated with the ilium because the combined width of the expanded ends of these ribs match approximately the length of the iliac blade (Fig. 2c, d). However, in contrast to terrestrial amniotes, the sacral ribs are not fused to the sacral vertebrae, and their distal ends are thin, with only

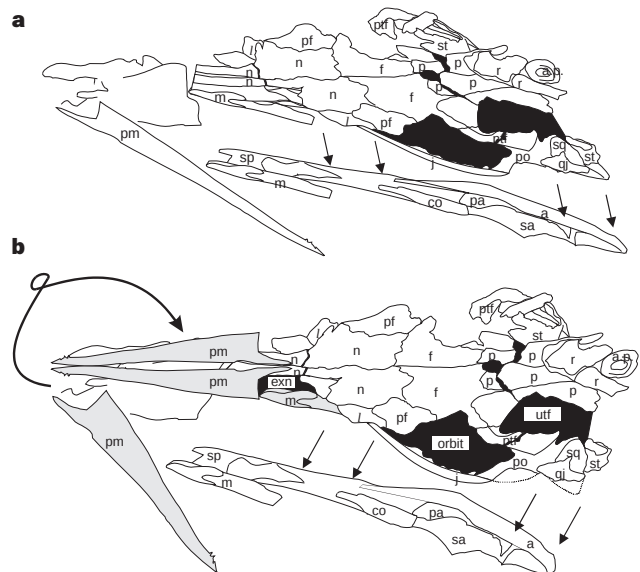
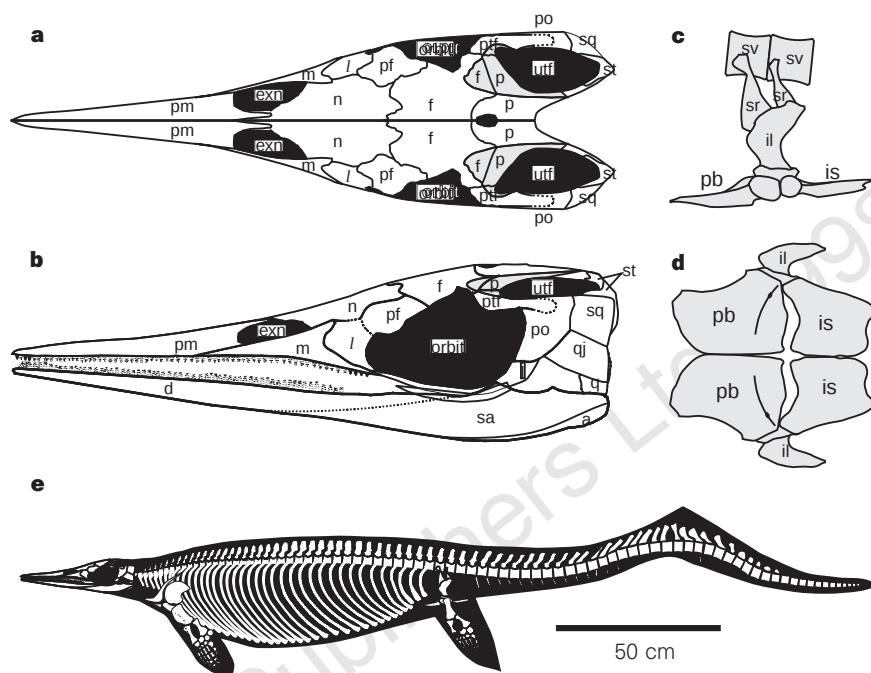


Figure 1 Retrodeformation of the skull of *Utatsusaurus hataii* (UHR 30691). **a**, Dorsal view of the skull before retrodeformation. The mandible, hidden beneath the skull, has been moved in the image to a parallel position. **b**, Same specimen after retrodeformation. Two disarticulated bones (premaxilla and maxilla, in grey) are placed back in their articulated positions in this image. Abbreviations: a, angular; a.p., atlantal pleurocentrum; co, coronoid; exn, external naris; f, frontal; j, jugal; l, lacrimal; m, maxilla; n, nasal; p, parietal; pa, prearticular; pf, prefrontal; pm, premaxilla; po, postorbital; ptf, postfrontal; q, quadrate; qj, quadratojugal; r, rib fragments; sa, surangular; sp, splenial; sq, squamosal; st, supratemporal; utf, upper temporal fenestra.

Figure 2 Reconstruction of *Utatusaurus hataii*. **a, b**, Skull based on retrodeformation of the new specimens. The height/width ratio was obtained from a three-dimensionally preserved skull of another basal ichthyosaurus (*Grippia longirostris*). **a**, Left lateral and **b**, dorsal views. **c, d**, Pelvic girdle based on new specimens. The girdle is attached to the vertebral column by two well-defined sacral ribs, but the attachment is not robust. **c**, Left lateral and **d**, ventral views. **e**, The entire skeleton based on the type materials¹⁴ and the new specimens. Elements from various specimens were resized according to the closest vertebral centra. Abbreviations: il, ilium; is, ischium; pb, pubis; sr, sacral rib; sv, sacral vertebra. See Fig. 1 for abbreviations for the skull bones.



slight vertical expansion. Therefore the attachment of the pelvic girdle to the vertebral column was probably not robust enough to support the body on land. Another transitional feature of *Utatusaurus* is the equal length of the humerus and the femur. In all other ichthyosaurs, even *Chensaurus*, the humerus is longer than the femur, whereas in most terrestrial amniotes the femur is longer than the humerus¹⁶. The hindlimb is incompletely characterized in *Utatusaurus* (Fig. 2e), but it seems to be larger than the forelimb. The reduction of the hindlimb is a common aquatic adaptation³.

The new specimens refute the recent proposal that the ichthyosaurian upper temporal fenestra is of non-diapsid origin¹⁷. The postfrontal in *Utatusaurus* has a long posterior process which overlies the postorbital, but does not exclude the latter from the border of the upper temporal fenestra (Figs 1 and 2a, b). Thus the postorbital and squamosal meet lateral to the upper temporal fenestra, which is a primitive condition for diapsids that has never been established before for ichthyosaurs. In derived ichthyosaurs,

these two bones are completely eliminated from the fenestral margin by the postfrontal and supratemporal, creating the so-called 'parapsid' (ref. 18) condition, which can now be interpreted as a modified diapsid condition.

Problems in identifying the bones in the cheek region of ichthyosaurs have impaired analysis of ichthyosaurian relationships. One of the new specimens (UHR 30691) clearly shows three bones in the posterior cheek region, identified as the supratemporal, squamosal and quadratojugal (Fig. 2a, b). It is generally accepted that only two bones are present in this region in ichthyosaurs, based on studies of two derived ichthyosaurs^{19,20}. However, later studies show that other derived ichthyosaurs actually had three elements^{17,21,22}, as we found in *Utatusaurus*. The arrangement of the three bones in the new specimen suggests that the bone formerly identified as the squamosal in derived ichthyosaurs^{19,20} is actually the supratemporal, and that the squamosal is absent in some forms. The supratemporals of derived ichthyosaurs are large, unlike those in most other

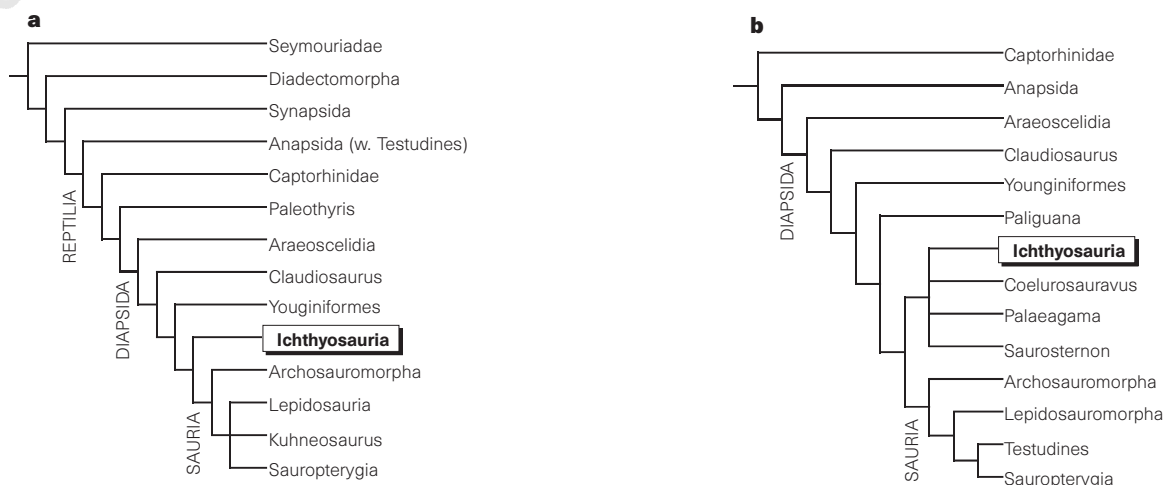


Figure 3 Phylogenetic position of the Ichthyosauria, according to reanalyses of two published data matrices. **a**, Strict consensus of two shortest trees (794 steps, CI 0.49) obtained from the data matrix of Rieppel and DeBraga⁹ by adding ichthyosaurs. **b**, Strict consensus of 30 shortest trees (391 steps, CI 0.58) obtained

from the data matrix of Caldwell¹⁰ by recoding ichthyosaurian characters. In both cases, the Ichthyosauria appeared as the sister group of the Sauria. The addition of ichthyosaurs did not change the position of turtles in either case. Both trees were simplified by using inclusive clade names. See Methods for details.

vertebrates except in some snakes, and this was one of the reasons why the bone was previously identified as the squamosal. The supratemporal of *Utatusaurus*, however, is of moderate size, so this bone seems to have progressively increased in size in the Ichthyosauria.

Most major amniote groups, and even some non-amniote tetrapods, have at some time been proposed to be the sister group of ichthyosaurs⁵. Although a consensus is emerging that ichthyosaurs are diapsids^{5,6,10,11,23}, there is still opposition^{17,24}, largely because of the lack of information about the basal ichthyosaurs^{6,17}. Two recent cladistic studies concluded that the Ichthyosauria is the sister group of the Sauropterygia (which includes plesiosaurs and placodonts), which together form the sister group of the archosauromorphs^{10,11}. However, only limited data from basal ichthyosaurs were available for these studies^{10,11}, and re-examination of one of the data matrices¹⁰ revealed that about a third of the characters would be coded differently if *Utatusaurus* were included. Moreover, this placement of sauropterygians disagrees with the more widely accepted hypothesis, in which sauropterygians are more closely related to lepidosaurs than to archosaurs^{8,9}. To test these competing hypotheses, we reanalysed two data matrices^{9,10} by incorporating the data obtained from the new specimens of *Utatusaurus*. In both cases (Fig. 3), ichthyosaurs appeared within the Diapsida but outside the Sauria, whereas sauropterygians were separated from ichthyosaurs and formed a clade with lepidosauromorphs. The new specimens of *Utatusaurus* confirm once again the importance of basal taxa to the assignments of character polarities and to the outcome of phylogenetic analyses²⁵. □

Methods

Retrodeformation. The effects of tectonic deformation on taxonomy are important in invertebrate palaeontology^{26,27}. Information available for the retrodeformation of a vertebrate fossil is usually limited, but the original shape can be restored with as few as two pairs of measurements⁷. The dorsal view of the skull, which is not parallel to the bedding plane, was retrodeformed according to the measurements from the skull roof itself, as described⁷. The image of the entire skeleton along the bedding plane was retrodeformed, based on measurements from the vertebral centra made according to ref. 28 and confirmed as described in ref. 7. In both cases, the bilateral symmetry of the body was well restored.

Phylogenetic analyses. The first phylogenetic hypothesis (Fig. 3a) was derived by adding the Ichthyosauria to the data matrix of Rieppel and DeBraga⁹ (see Supplementary Information). Several character codings were amended according to ref. 29. In addition, the following characters were modified: 30, there is no supratemporal in sauropterygians (recoded); 33, there is no posterior process on the jugal in basal ichthyosaurs (new character state added); 83, the coronoid process is present on the surangular in ichthyosaurs (new character state added). Two characters, namely (162) short limbs and (163) short manus/pes, were omitted because limb, manus and pes each comprise multiple elements, and the whole structure may be shortened in many different ways; and because such shortenings are also known in other aquatic tetrapods with different origins, such as cetaceans and mosasaurs, so they probably reflect aquatic adaptation rather than a common origin. The second phylogenetic hypothesis (Fig. 3b) was derived by recoding ichthyosaurs in the data matrix of Caldwell¹⁰. Only ichthyosaurian characters were recoded in the modified character matrix (see Supplementary Information). In both cases, the ichthyosaurian characters were coded according to *Utatusaurus*; in the absence of information for *Utatusaurus*, references were made to successively more derived ichthyosaurs, namely *Chensaurus*, *Cymbospondylus* and *Ichthyosaurus*. All phylogenetic analyses were made by using the heuristic search option of PAUP 3.1.1 (100 random addition sequences for both TBR and SPR)³⁰.

Received 12 November 1997; accepted 23 February 1998.

- McGowan, C. *Dinosaurs, Spittfires, and Sea Dragons* (Harvard Univ. Press, Cambridge and London, 1991).
- Motani, R., You, H. & McGowan, C. Eel-like swimming in the earliest ichthyosaurs. *Nature* **382**, 347–348 (1996).

- Carroll, R. L. Evolutionary constraints in aquatic diapsid reptiles. *Spec. Pap. Palaeontol.* **33**, 145–155 (1985).
- Carroll, R. L. & Dong, Z. M. *Hupehsuchus*, an enigmatic aquatic reptile from the Triassic of China, and the problem of establishing relationships. *Phil. Trans. R. Soc. Lond. B* **331**, 131–153 (1991).
- Callaway, J. M. *Systematics, Phylogeny, and Ancestry of Triassic Ichthyosaurs* (Reptilia, Ichthyosauria) (thesis, Univ. Rochester, 1989).
- Massare, J. A. & Callaway, J. M. The affinities and ecology of Triassic ichthyosaurs. *Geol. Soc. Am. Bull.* **102**, 409–416 (1990).
- Motani, R. New technique for retrodeforming tectonically deformed fossils, with an example for ichthyosaur specimens. *Lethaia* **30**, 221–228 (1997).
- Rieppel, O. Osteology of *Simosaurus gaillardoti* and the relationships of stem-group Sauropterygia. *Fieldiana Geol.* **28**, 1–85 (1994).
- Rieppel, O. & DeBraga, M. Turtles as diapsid reptiles. *Nature* **384**, 453–454 (1996).
- Caldwell, M. W. Ichthyosauria: A preliminary phylogenetic analysis of diapsid affinities. *N. Jb. Geol. Paläontol. Abh.* **200**, 361–386 (1996).
- Merck, J. W. A phylogenetic analysis of the euryapsid reptiles. *J. Vert. Paleontol.* **17** (suppl.), 65A (1997).
- Minoura, N. et al. Excavation of Early Triassic ichthyosaur fossils from Ogatsu, Miyagi Pref., Japan. *Geol. Stud.* **42**, 215–232 (1993).
- Minoura, N. in *Paleoasian Ocean to Paleo-Pacific Ocean* 64–68 (Hokkaido Univ., Sapporo, 1994).
- Shikama, T., Kamei, T. & Murata, M. Early Triassic Ichthyosaurus, *Utatusaurus hataii* gen. et sp. nov., from the Kitakami Massif, Northeast Japan. *Sci. Rep. Tohoku Univ., Sendai, Second Ser. (Geol.)* **48**, 77–97 (1978).
- Motani, R. Phylogeny of the Ichthyosauria (Amniota: Reptilia) with special reference to Triassic forms. *J. Vert. Paleontol.* **17** (suppl.), 66A (1997).
- Romer, A. S. *Osteology of the Reptiles* (Univ. Chicago Press, Chicago and London, 1956).
- Maisch, M. W. A case against a diapsid origin of the Ichthyosauria. *N. Jb. Geol. Paläontol. Abh.* **205**, 111–127 (1997).
- Williston, S. W. *Osteology of the Reptiles* (Harvard Univ. Press, Cambridge, MA, 1925).
- McGowan, C. The cranial morphology of the Lower Liassic latipinnate ichthyosaurs of England. *Bull. Br. Mus. Nat. Hist. Geol.* **24**, 1–109 (1973).
- Romer, A. S. An ichthyosaur skull from the Cretaceous of Wyoming. *Contrib. Geol. Univ. Wyo.* **7**, 27–41 (1968).
- Kirtan, A. M. *A Review of British Upper Jurassic Ichthyosaurs* Thesis (Univ. Newcastle, 1983).
- Godefroit, P. The skull of *Stenopterygius longifrons* (Owen, 1881). *Rev. Paléobiol., Genève, Vol. Spéc.* **7**, 67–84 (1993).
- Tarsitano, S. A model for the origin of ichthyosaurs. *N. Jb. Geol. Paläontol. Abh.* **164**, 143–145 (1982).
- Riess, J. Fortbewegungsweise, Schwimmbiophysik, und Phylogenie der Ichthyosaurier. *Palaeontogr. Abt. A* **192**, 93–155 (1986).
- Gauthier, J., Kluge, A. G. & Rowe, T. Amniote phylogeny and the importance of fossils. *Cladistics* **4**, 105–209 (1988).
- Cooper, R. A. Interpretation of tectonically deformed fossils. *N. Z. J. Geol. Geophys.* **33**, 321–332 (1990).
- Hughes, N. C. & Jell, P. A. A statistical/computer graphic technique for assessing variation in tectonically deformed fossils and its application to Cambrian trilobites from Kashmir. *Lethaia* **25**, 317–330 (1992).
- Ando, T., Kawamura, M. & Minoura, N. Restoration of tectonically deformed Early Triassic ichthyosaurs. *Abstr. Progr. 1997 Annu. Meet. Geol. Soc. Japan* **160** (1997).
- Lee, M. Reptile relationships turn turtle. *Nature* **389**, 245–246 (1997).
- Swofford, D. L. *PAUP 3.1.1* (Smithsonian Inst., Washington DC, 1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

Acknowledgements. We thank C. McGowan for his support; K. Padian for reading the manuscript; M. Caldwell, A. Hungerbühler, M. Kawamura, J. Merck and H.-D. Sues for discussions; Y. Tomida and M. Manabe for support at NSM; T. Kato, T. Kuwajima, G. Kawakami, H. Normura, D. Suzuki, M. Takahashi for technical support; and the Fujiwara Natural History Foundation, Tokyo, and Miller Institute for Basic Research in Science, Berkeley (R.M.) and the Fukuda Geological Institute (N.M.) for financial support.

Correspondence and requests for materials should be addressed to R.M. (e-mail: motani@ucmp1.berkeley.edu).

Dynamics of North American breeding bird populations

Timothy H. Keitt* & H. Eugene Stanley†

* Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA and National Center for Ecological Analysis and Synthesis, University of California Santa Barbara, 735 State Street, Suite 300, Santa Barbara, California 93101, USA

† Center for Polymer Studies and Department of Physics, Boston University, Boston 02215, USA

Population biologists have long been interested in the variability of natural populations^{1–6}. One approach to dealing with ecological complexity is to reduce the system to one or a few species, for which meaningful equations can be solved. Here we explore an alternative approach^{7,8} by studying the statistical properties of a data set containing over 600 species, namely the North American breeding bird survey⁹. The survey has recorded annual species abundances over a 31-year period along more than 3,000 observation routes¹⁰. We now analyse the dynamics of population

variability using this data set, and find scaling features in common with inanimate systems composed of strongly interacting subunits¹¹. Specifically, we find that the distribution of changes in population abundance over a one-year interval is remarkably symmetrical, with long tails extending over six orders of magnitude. The variance of the population over a time series increases as a power-law with increasing time lag, indicating long-range correlation in population size fluctuations¹². We also find that the distribution of species lifetimes (the time between colonization and local extinction) within local patches is a power-law with an exponential cutoff imposed by the finite length of the time series. Our results provide a quantitative basis for modelling the dynamics of large species assemblages.

For each of the approximately 2,500 routes surveyed in a given year, the North American Breeding Bird Survey (NA BBS) gives $N_s(t)$, the number of birds of a given species s detected on a given route in year t . The challenge is to extract information relevant to population variability from these data.

We are interested in the dynamics, so we must compare values of $N_s(t)$ in successive years. We form the quantities $R_s(t) \equiv N_s(t+1)/N_s(t)$, which give the rate of increase or decrease of each species on each route. As population growth is multiplicative, the distribution of abundance $N_s(t)$ of a species through time is often approximately log-normal^{13–15}. The ratio of two numbers, each drawn from a log-normal distribution, is also log-normally distributed, so we might expect that the distribution of $R_s(t)$ would be log-normal as well. We find that the distribution of $R_s(t)$ is not log-normal, but is instead power-law, with the tails separated by six orders of magnitude (Fig. 1a), such that

$$P(R_s) \propto \begin{cases} R_s^\alpha & \text{if } R_s \leq 1 \\ R_s^{-\alpha} & \text{if } R_s \geq 1 \end{cases} \quad (1)$$

where the exponent $\alpha = 2$. Thus, for the species considered here, there is no characteristic scale of fluctuation in population size, but instead we find a broad spectrum of growth rates.

We may also consider the more commonly used^{16,17} logarithm of the ratio of successive abundances, $r_s \equiv \log R_s$. Equation (1) then becomes $P(r_s) \propto \exp(-\alpha|r_s|)$.

What is remarkable about the results shown in Fig. 1a is that, despite the inclusion of a large number of species, the overall

distribution can be described by relatively simple mathematical equations. The distribution is highly symmetrical, so exactly as many species are increasing in abundance as are decreasing.

In this analysis, the quantities $R_s(t)$ and $r_s(t)$ measure fluctuations in population size over a one-year interval. It is also possible to analyse the data set for an arbitrary time lag Δt , where Δt can take any value from 1 to 30. We might expect that the variance of the distribution functions corresponding to Fig. 1a will increase with the magnitude of Δt (refs 18, 19). We do indeed find an increase (Fig. 1b), and moreover this increase is a power law of the form $(\Delta t)^{2H}$, where H is the Hurst exponent. For a random walk, successive changes are uncorrelated. The variance of a random walk increases at a particular rate $H = 0.5$. Complex dynamics generate serial correlations over time so that the variance increases more rapidly ($H > 0.5$) or more slowly ($H < 0.5$) than for a random walk. For the NA BBS data, we find $H = 0.14$, a value considerably smaller than 0.5, which implies the existence of long-range correlations. Correlation in population fluctuations could also arise from correlations in climate variables that influence reproduction and mortality²⁰ (for example, the severe North American winter of 1975–1976 is thought to have caused large declines in the numbers of some resident bird species^{21–23}).

Fluctuations in population size can lead to local extinctions^{24,25} and turnover in local species composition²⁶. For example, survivorship of local spider populations on Caribbean islands followed a hollow curve that deviated from exponential²⁷. Differences in survivorship among species were attributed to different species-specific life-history strategies. Local extinction dynamics can also be studied quantitatively by considering the distribution of 'lifetimes' of species within a local region²⁸. We define the lifetime of a species as $\tau \equiv t_e - t_c$, where t_c denotes the year in which the species colonized a patch (the first year that the species was recorded on a given route), and t_e denotes the year that the species became locally extinct (the last year that the species was recorded on that route). The distribution of species' lifetimes for the NA BBS data follows a Yule distribution described by a power law, modified by an exponential cutoff (Fig. 2a)

$$P(\tau) = A\tau^{-\alpha}e^{-\tau/\tau_{ch}} \quad (2)$$

where τ_{ch} sets the timescale at which the power-law scaling no longer

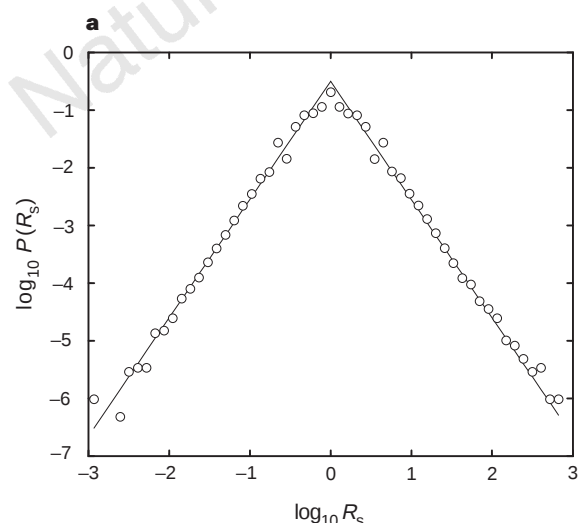
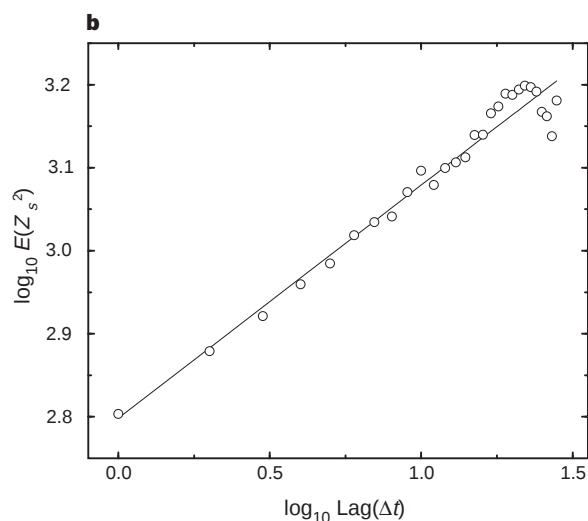


Figure 1 Scaling of population growth rates in the NA BBS data set. **a**, Distribution of population growth rates $R_s \equiv N_s(t+1)/N_s(t)$ across all species in the data. The growth rate R_s is calculated by dividing species abundances in successive years. Abundances are taken as the total number of individuals of a particular species counted within each survey route. The histogram values at the centre of the distribution are probably influenced by Poisson sample error present in small counts. In a separate analysis, we exclude small counts and find that the tails of



the distribution are unaffected. **b**, Characterization of fractal scaling in population dynamics. The variance of the fluctuation distribution $E(Z_s^2)$ increases as a power-law function of time lag Δt , where $Z_s \equiv N_s(t+1) - N_s(t)$. Because the number of points separated by a given lag Δt decreases rapidly with increasing lag, deviations from power-law scaling are expected for lags approaching the 31-year length of the time series. The slope of 0.28 corresponds to a Hurst exponent of 0.14.

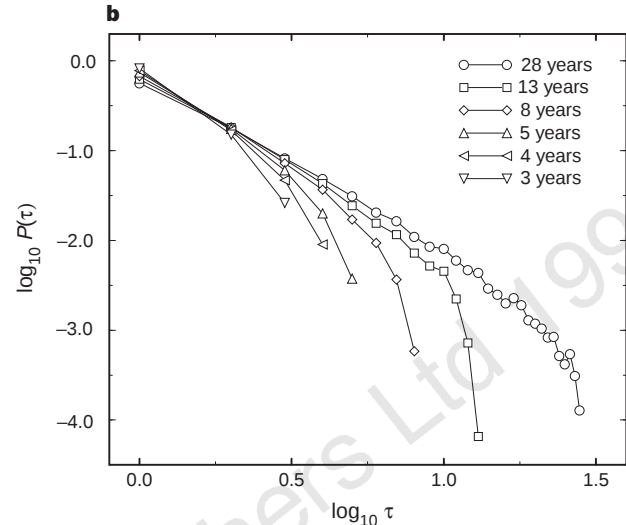
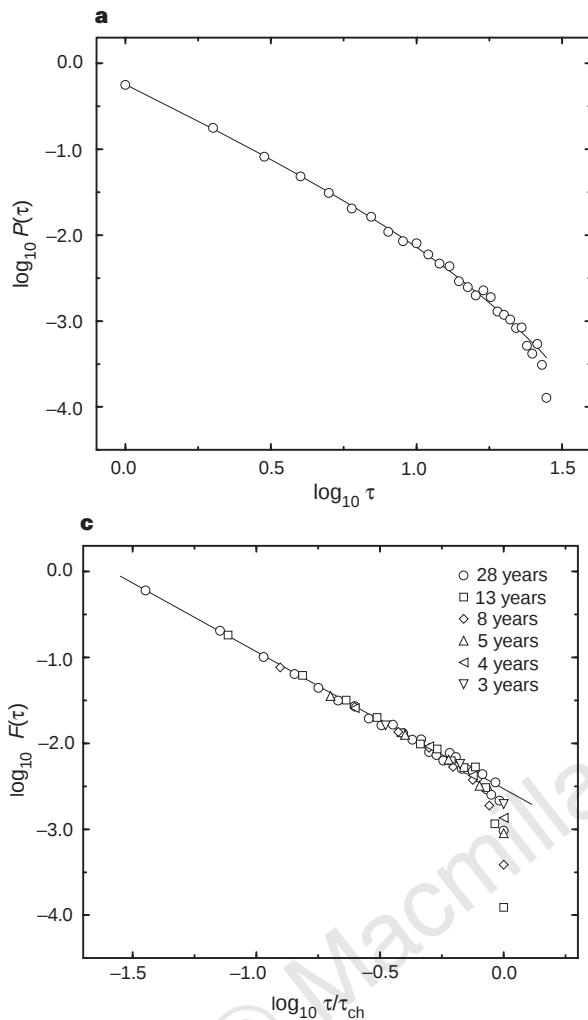


Figure 2 Scaling of local species' lifetimes in the NA BBS data set. **a**, Frequency distribution of species lifetimes within local patches for all species. Parameters of the data fit are obtained by weighted least-squares regression. The sum of the distribution is normalized to unity. The lifetime of a species within a patch is the time between colonization and local extinction. Colonization occurs when a species is recorded, but was absent the previous year; extinction occurs when a species is absent, but was recorded the previous year. Time series not surveyed in each year between colonization and extinction are not included, as well as time series that begin or end in the first or last year of the survey data. **b**, Finite size analysis of the species' lifetime distribution. The distributions are from non-overlapping subsets of each time series, broken into increasingly shorter sequences. The number of years shown are the maximum possible species' lifetimes between the first and last year of the individual time series. **c**, Test of generality of species' lifetimes distribution. When axes are rescaled to remove finite scaling effects, the data collapse onto a single power-law curve. For each finite subset of the data, we compute $F(\tau) = P(\tau)/A\tau_{ch}^{-\alpha}e^{-\tau/\tau_{ch}}$ and plot $F(\tau)$ against the rescaled time axis τ/τ_{ch} .

holds owing to the finite size of the data set. We find that $\alpha = 1.61 \pm 0.02$ and $\tau_{ch} = 14$ years for the NA BBS data. The parameter A is an arbitrary normalization constant.

The exponential term in equation (2) introduces a characteristic timescale of $\tau_{ch} = 14$ years into the distribution of extinction times. We can test whether the exponential term is due to finite size effects by plotting a series of lifetime distributions, each computed from non-overlapping subsets of the time series associated with each survey route (Fig. 2b). The resulting distributions are increasingly truncated by the lengths of the time series. When plotted on rescaled axes (Fig. 2c), the distributions collapse onto a single power-law curve. Thus, we find that the power-law distribution is unaffected by shortening the time series; only the exponential term is affected, suggesting that the power-law behaviour extends beyond the 31-year extent of this data set.

The presence of scaling, both in population variability (Fig. 1a) and in local species' lifetimes (Fig. 2c), may have implications for understanding population dynamics in general. Scaling of system variables is often observed in physical systems that exhibit cooperative behaviour^{29,30}. Scaling arises in these inanimate systems because each particle interacts directly with a few neighbouring particles and as these neighbouring particles interact with their neighbours, interactions can 'propagate' long distances, resulting in power-law distributions. Similarly, species in an ecosystem interact directly with some (but not all) other species, which in turn interact with other species so that interactions can 'propagate'.

Interactions can propagate not only among species in an ecosystem, but through space as well. In both island biogeography³¹ and metapopulation theory²⁴, the probability of a patch (island) being

occupied is often modelled by the incidence function $P(x_i = 1) = c_i/c_i + e_i$, where $x_i = 1$ if the patch is occupied, c_i is the colonization probability, and e_i is the extinction probability^{26,32}. On islands, where distances between populations are large, we expect that the species lifetime distribution should decay exponentially, because the probability $P(x_i = 1)$ is roughly constant. However, metapopulations are often more strongly coupled than island populations, and can experience a much stronger 'rescue effect'³³ whereby local populations are saved from extinction by immigration from nearby patches. This rescue effect may explain the long-tailed distribution of species lifetimes. □

Received 23 December 1997; accepted 19 February 1998.

- May, R. M. Biological populations with nonoverlapping generations: Stable points, stable cycles, and chaos. *Science* **186**, 645–647 (1974).
- Ives, A. R. Chaos in time and space. *Nature* **353**, 214–215 (1991).
- Turchin, P. Rarity of density dependence or population regulation with lags? *Nature* **344**, 660–663 (1990).
- Arino, A. & Pimm, S. L. On the nature of population extremes. *Evol. Ecol.* **9**, 429–443 (1995).
- Rhodes, C. J. & Anderson, R. M. Power laws governing epidemics in isolated populations. *Nature* **381**, 600–602 (1996).
- Curnutt, J. L., Pimm, S. L. & Maurer, B. A. Population variability of sparrows in space and time. *Oikos* **76**, 131–144 (1996).
- Brown, J. H. *Macroecology* (Univ. of Chicago Press, Chicago, 1995).
- Keitt, T. H. & Marquet, P. A. The introduced Hawaiian avifauna reconsidered: Evidence for self-organized criticality? *J. Theor. Biol.* **182**, 161–167 (1996).
- Peterjohn, B. G. The North American breeding bird survey. *J. Am. Birding Assoc.* **26**, 386–398 (1994).
- Sauer, J. R., Hines, J. E., Gough, G., Thomas, I. & Peterjohn, B. G. *The North American breeding bird survey results and analysis, version 96.3* (Patuxent Wildlife Research Center, Laurel, MD, 1997). (<http://www.mbr.nbs.gov/bbs/bbs.html>)
- Stanley, M. H. R. et al. Scaling behaviour in the growth of companies. *Nature* **379**, 804–806 (1996).
- Beran, J. *Statistics for Long Memory Processes* (Chapman and Hall, New York, 1994).
- Preston, F. W. The canonical distribution of commonness and rarity: I. *Ecology* **43**, 185–215 (1962).
- MacArthur, R. H. On the relative abundance of bird species. *Proc. Natl Acad. Sci. USA* **43**, 293–295 (1957).
- Sugihara, G. Minimal community structure: An explanation of species abundance. *Am. Nat.* **116**, 770–787 (1980).

16. Hassell, M. P., Lawton, J. H. & May, R. M. Patterns of dynamical behavior in single-species populations. *J. Anim. Ecol.* **45**, 471–486 (1976).
17. Royama, T. *Analytical Population Dynamics* (Chapman and Hall, London, 1992).
18. Connell, J. H. & Sousa, W. P. On the evidence needed to judge ecological stability or persistence. *Am. Nat.* **121**, 789–824 (1983).
19. Pimm, S. L. & Redfearn, A. The variability of population densities. *Nature* **334**, 613–614 (1988).
20. Steele, J. H. A comparison of terrestrial and marine ecological systems. *Nature* **313**, 355–358 (1985).
21. Pitts, T. D. Eastern bluebird populations fluctuations in Tennessee during 1970–1979. *The Migrant* **52**, 29–37 (1981).
22. Sauer, J. R. & Droege, S. Recent population trends of the eastern bluebird. *Wilson Bull.* **102**, 239–252 (1990).
23. Sauer, J. R., Pendleton, G. W. & Peterjohn, B. G. Evaluating causes of population change in North American insectivorous songbirds. *Conservation Biol.* **10**, 465–478 (1996).
24. Hanski, I. & Gilpin, M. Metapopulation dynamics: Brief history and conceptual domain. *Biol. J. Linn. Soc.* **42**, 3–16 (1991).
25. Hanski, I., Foley, P. & Hassell, M. Random walks in a metapopulation: How much density dependence is necessary for long-term persistence? *J. Anim. Ecol.* **65**, 274–282 (1996).
26. Diamond, J. M. & May, R. M. Species turnover rates on islands: Dependence on census interval. *Science* **197**, 266–270 (1977).
27. Schoener, T. W. & Spiller, D. A. High population persistence in a system with high turnover. *Nature* **330**, 474–477 (1987).
28. Pimm, S. L., Diamond, J., Reed, T. M., Russell, G. J. & Verner, J. Times to extinction for small populations of large birds. *Proc. Natl. Acad. Sci. USA* **90**, 10871–10875 (1993).
29. Stanley, H. E. *Introduction to Phase Transitions and Critical Phenomena* (Oxford Univ. Press, Oxford, 1971).
30. Stanley, H. E. Power laws and universality. *Nature* **378**, 554 (1995).
31. MacArthur, R. H. & Wilson, E. O. *Island Biogeography* (Princeton Univ. Press, 1967).
32. Hanski, I. A practical model of metapopulation dynamics. *J. Anim. Ecol.* **63**, 151–162 (1994).
33. Brown, J. H. & Kodric-Brown, A. Turnover rates in insular biogeography: effect of immigration on extinction. *Ecol.* **58**, 445–449 (1977).

Acknowledgements. This work was partially supported by the Santa Fe Institute, Thaw Foundation and National Science Foundation. We thank B. Peterjohn and the Patuxent Wildlife Research Center, United States Department of the Interior, for providing the NA BBS data in digital form, and L. A. N. Amaral, P. Bak, J. H. Brown, P. A. Marquet, B. Maurer, R. M. May, B. T. Milne, M. Paczuski, B. Peterjohn, S. L. Pimm, R. V. Solé and M. Taper for their comments.

Correspondence and requests for materials should be addressed to T.H.K. (e-mail: keitt@nceas.ucsb.edu).

Herbicide resistance caused by spontaneous mutation of the cytoskeletal protein tubulin

Richard G. Anthony, Teresa R. Waldin, John A. Ray*, Simon W. J. Bright* & Patrick J. Hussey

School of Biological Sciences, Royal Holloway University of London, Egham Hill, Egham, Surrey TW20 0EX, UK

* Zeneca Agrochemicals, Jeolott's Hill Research Station, Bracknell, Berkshire RG42 6ET, UK

The dinitroaniline herbicides (such as trifluralin and oryzalin) have been developed for the selective control of weeds in arable crops. However, prolonged use of these chemicals has resulted in the selection of resistant biotypes of goosegrass, a major weed. These herbicides bind to the plant tubulin protein but not to mammalian tubulin¹. Here we show that the major α -tubulin gene of the resistant biotype has three base changes within the coding sequence. These base changes swap cytosine and thymine, most likely as the result of the spontaneous deamination of methylated cytosine. One of these base changes causes an amino-acid change in the protein: normal threonine at position 239 is changed to isoleucine. This position is close to the site of interaction between tubulin dimers in the microtubule protofilament. We show that the mutated gene is the cause of the herbicide resistance by using it to transform maize and confer resistance to dinitroaniline herbicides. Our results provide a molecular explanation for the resistance of goosegrass to dinitroaniline herbicides, a phenomenon that has arisen, and been selected for, as a result of repeated exposure to this class of herbicide.

The dinitroaniline herbicides disrupt meristem development in the roots and shoots as a result of the net depolymerization of cellular microtubules^{2,3}. The repeated use of the dinitroaniline herbicides on the cotton and soybean fields in North America has resulted in the appearance of biotypes of goosegrass, *Eleusine indica* (Fig. 1), that have evolved resistance to the herbicides^{4–6}. Dose–

Table 1 Differential sensitivity of the sensitive and resistant biotypes to dinitroaniline herbicides and to the structurally unrelated pronamide herbicide

Herbicide	ID ₅₀ (mg l ⁻¹)		R/S*
	Sensitive	Resistant	
Trifluralin	0.04	1.70	42.5
Oryzalin	0.01	0.60	60.0
Pronamide	0.55	0.59	1.1

The ID₅₀ value was calculated from seedling dose–response curves.

* Ratio of ID₅₀ for resistant biotype to ID₅₀ for susceptible biotype.

response studies in the laboratory⁶ showed that a dinitroaniline-resistant biotype tolerated up to 60-fold higher concentrations of herbicide than did the dinitroaniline-sensitive biotype, but that both biotypes were equally sensitive to the structurally unrelated antimicrotubule herbicide pronamide⁷ (Table 1). It is reasonable to propose, in light of the effect of dinitroanilines on plant microtubules^{1,8}, that a mutation has arisen in a microtubule protein. In a preliminary investigation we showed that the major α -tubulin isotype in the resistant biotype had an altered electrophoretic mobility on two-dimensional gels compared to the isotype of the sensitive biotype. In further experiments we focused on a molecular analysis of the α -tubulin genes.

We constructed two complementary DNA libraries in UniZAP vectors (Stratagene), using poly(A)⁺ RNA isolated from seedlings derived from selfed sensitive and resistant biotypes. The sensitive-biotype library was probed with radiolabelled *Zmtua1*, a maize α -tubulin cDNA clone⁹. We obtained 13 clones that hybridized to the maize α -tubulin probe. DNA sequencing revealed that 11 of the 13 cDNA clones had identical 3' non-coding regions, indicating that these represented the most prevalent α -tubulin transcript. This group included the full-length cDNA clone *EiStua1*. The remaining two clones were distinct from the prevalent cDNA and from each

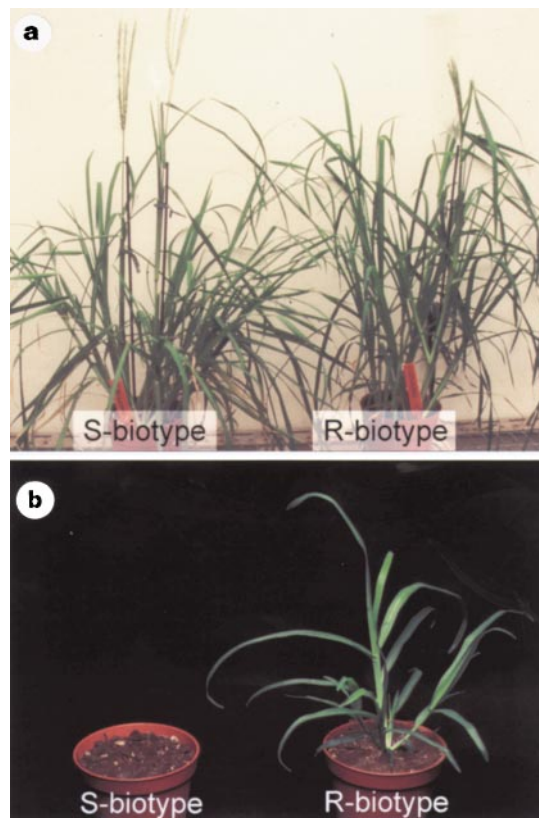


Figure 1 Dinitroaniline-sensitive (S) and -resistant (R) biotypes of *Eleusine indica* (goosegrass). **a**, S and R biotypes grown to maturity in the absence of herbicide. **b**, S and R biotypes grown on a discriminatory dose of trifluralin (1 mg l⁻¹).

EiStua1

cDNA/genomic DNA 715
 Protein -V--I--S--S--L--T--A--S--L--R--F--
 239

EiRtua1

cDNA/genomic DNA 715
 Protein -V--I--S--S--L--I--A--S--L--R--F--
 239

other. To isolate a full-length homologue of *EiStua1* from the resistant-biotype library, the resistant-biotype library was hybridized with a 94-base-pair (bp) fragment derived from the amplification of the 5' non-coding region of *EiStua1* using the polymerase chain reaction (PCR). Clones hybridizing to this probe accounted for ~80% of all plaques that hybridized to the maize α -tubulin clone using the same filters. Thus the prevalence of the *EiStua1* cDNA observed in the sensitive-biotype library was reproduced in the resistant-biotype library. Three cDNA clones from the resistant biotype were isolated as homologues of *EiStua1*, of which one, *EiRtua1*, contained a full-length insert. Sequence analysis of both the 5' and the 3' non-coding regions indicated that this cDNA was equivalent to *EiStua1*.

We sequenced full-length clones *EiStua1* and *EiRtua1* completely and compared the nucleotide sequence and encoded amino-acid sequence. Five bases differ between *EiStua1* and *EiRtua1*, including two in the 5' non-coding region (T9 $\rightarrow$ C and A8 $\rightarrow$ C) and three in the coding region (T524 $\rightarrow$ C, C715 $\rightarrow$ T and T989 $\rightarrow$ C). Four of the changes exchange cytosine and thymine; these changes may have arisen as a result of deamination of methylated cytosine in one of the lines. The C715 $\rightarrow$ T base change in the coding sequence of *EiRtua1* results in the amino-acid substitution of Thr 239 with Ile 239; this is a replacement of a polar with a nonpolar residue (Fig. 2). The T524 $\rightarrow$ C and T989 $\rightarrow$ C changes are same-sense changes; that is, they do not alter the amino-acid sequence.

EiRtua1 was used directly to test whether it could confer herbicide resistance in transgenic plant cells. We made several constructs (Fig. 3) for transformation of maize Black Mexican Sweetcorn (BMS) suspension culture cells using silicon carbide fibres. We used the 35S promoter (Fig. 3) to direct the expression of the tubulin genes. The selection of transformed clones was on bialaphos, resistance being conferred to the transformants by expression of the *bar* gene which was also under the control of the 35S promoter. To ensure cell viability, it was essential that a β -tubulin gene (*Zmtub2*; ref. 10) was expressed with either *EiStua1* (construct S (sensitive)) or *EiRtua1* (construct R (resistant)) in the same cell lines.

All of the clones that contained construct R, carrying the *EiRtua1/Zmtub2* coding sequence, were trifluralin resistant on a dose of herbicide (0.1 mg l⁻¹) that easily discriminated between these clones and those carrying construct S, containing the

Figure 2 A single base mutation results in an amino-acid difference between the goosegrass major α -tubulin gene from the sensitive biotype (*EiStua1*) and from the resistant biotype (*EiRtua1*). Segments of nucleic-acid sequence, and the corresponding deduced amino-acid sequence, from *EiStua1* and *EiRtua1* are shown.

EiStua1/Zmtub2 coding sequence. The clones varied in the number of copies of the inserted construct containing from one to ten copies. We tested three resistant clones (R1, possessing two copies of construct R, and R2 and R3, possessing a single copy of the same construct) for resistance to different dinitroanilines (trifluralin and oryzalin; Fig. 4) and to a structurally unrelated herbicide, pronamide⁷. These were compared directly with controls (wild-type BMS calli and a BMS clone containing a single copy of construct S) on the same plates (Fig. 4). The clones were plated onto gelled media containing increasing concentrations of the herbicides (0.0–1 mg l⁻¹). The controls did not grow on any of the herbicides, whereas the R1, R2 and R3 clones proliferated on both dinitroanilines. Some variation in the level of resistance was seen, particularly between R1 and the R2/R3 clones, and this is related to the higher copy number and thus the higher level of expression of the *EiRtua1* coding sequence in the R1 clone (data available on request). This implies that the level of resistance is dependent on the level of expressed resistant α -tubulin. All resistant clones and both controls were sensitive to pronamide, an antimicrotubule drug that acts differently to the dinitroanilines (Fig. 4).

To confirm that the resistant α -tubulin transgene product was incorporated into the microtubule arrays we made a further construct, R-tag, which incorporated epitope tags on both the α - and the β -tubulins (Fig. 3). The clones resulting from the BMS cells transformed with this construct were selected on bialaphos and trifluralin (0.1 mg l⁻¹) and a cell suspension generated with both herbicides incorporated into the liquid medium. The cells were prepared for immunofluorescence¹¹ and stained with anti-haemagglutinin antibodies to detect the resistant α -tubulin (Fig. 5a). The α -tubulin/haemagglutinin epitope tag is incorporated into cortical, spindle and phragmoplast arrays of microtubules. In addition, an immunoblot of a one-dimensional gel loaded with equal amounts of protein from either construct S-tag or R-tag clones, was probed with anti- α -tubulin antibodies (Fig. 5b). The results show that the α -tubulin/haemagglutinin epitope tag is more abundant than the endogenous tubulin. Similar results were obtained for β -tubulin/*c-myc*, using anti-*c-myc* antibodies to detect the β -tubulin in the microtubules, and anti- β -tubulin antibodies to probe the immunoblot (data available on request).

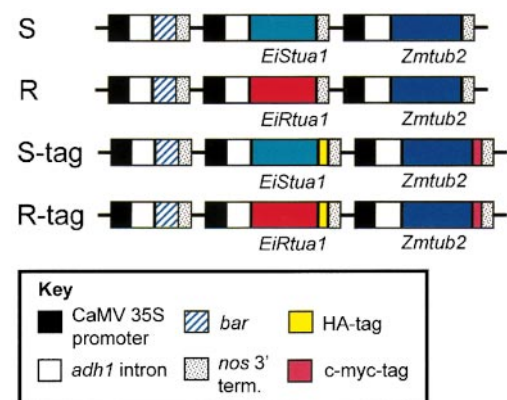


Figure 3 DNA cassettes containing α/β -tubulin genes fused to the hybrid promoter containing the cauliflower mosaic virus (CaMV) 35S promoter and the maize alcohol dehydrogenase (*adh1*) intron 1 (35S; ref. 17). These DNA cassettes were inserted into plasmid pJR1 (ref. 18) with the *bar* gene under the control of the 35S promoter for selection of transformants on bialaphos. *EiStua1* and *EiRtua1*, sensitive and resistant α -tubulin genes; *Zmtub2*, β -tubulin gene; *nos*, nopaline synthase gene; term, terminus.

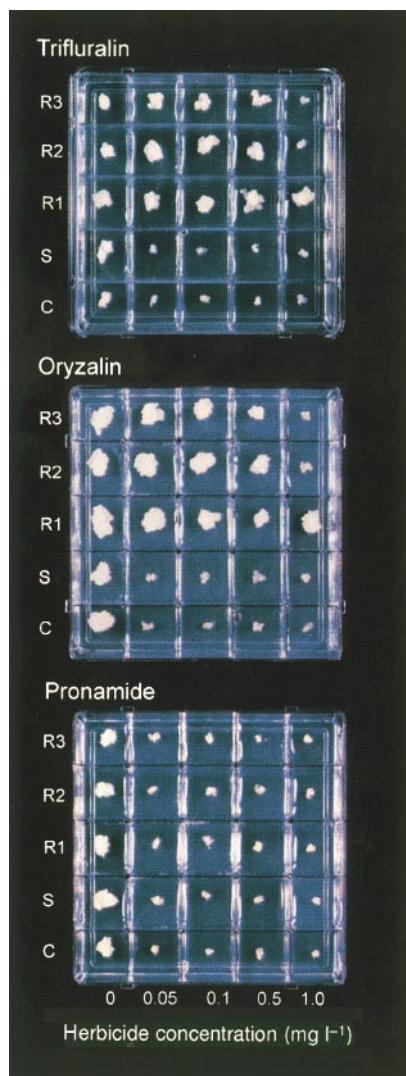


Figure 4 Response of control BMS calli and transformed BMS clones to dinitroaniline (trifluralin and oryzalin) herbicides, and to another antimicrotubule herbicide, pronamide, which has a different mode of action. Clones were subcultured onto Repli-plates containing BMS medium with herbicide concentrations up to 1.0 mg l^{-1} . The BMS calli (C) and the transgenic BMS clone expressing construct S, (S, single copy) are the controls. R1, R2 and R3 are transgenic BMS clones expressing construct R, (R1, two copies; R2/R3, single copies).

In general, the tubulin sequences from different organisms are highly conserved both within each tubulin subunit and between the different subunits (α , β and γ): amino-acid sequences of certain segments or single amino acids at specific positions of the polypeptide appear to be invariant¹². Threonine residue 239 is conserved in the deduced amino-acid sequences of all known α -tubulins, in the equivalent position in almost all β -tubulins (Thr 237; the few exceptions have a conservative serine substitution), and in all the known γ -tubulins (Thr 240), indicating that this is an important site for basic microtubule function. An atomic model of the $\alpha\beta$ -tubulin dimer has been determined by electron crystallography¹³. The crystal structure of FtsZ¹⁴, an evolutionarily distant relative of tubulin, is very similar and confirms that the structural fold of $\alpha\beta$ -tubulin as seen by electron microscopy is correct. The Thr 239 in α -tubulin is positioned at the end of the long central helix; thus it is close to the site that interacts with the β -monomer of the next dimer in the microtubule protofilament. In this situation, replacing threonine with isoleucine either disturbs the herbicide-binding site of tubulin so that it binds up to 60 times more weakly, or causes an increase in the stability of the dimer-dimer interaction.

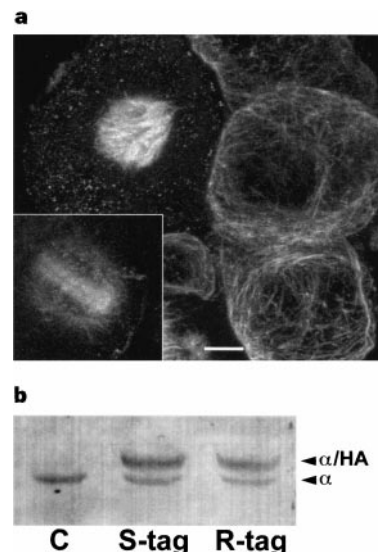


Figure 5 The α -tubulin gene product of resistant goosegrass biotypes is incorporated into the microtubule array. **a**, Localization of epitope-tagged resistant α -tubulin in transgenic BMS clones containing construct R-tag. Cortical, spindle and phragmoplast (inset) microtubule arrays are stained with anti-haemagglutinin antibodies (12CA5; ref. 19) to detect the resistant α -tubulin/haemagglutinin epitope tag. Bar, $10 \mu\text{m}$. **b**, Immunoblot of a one-dimensional gel loaded with total protein extracts from BMS calli (C) and of BMS clones containing a single copy of construct S-tag or R-tag. The α -tubulin/haemagglutinin epitope tag (α /HA) is the slower migrating band compared with the endogenous α -tubulin (α) in the S-tag and R-tag lanes.

We have discovered a molecular basis for resistance to dinitroaniline herbicides in goosegrass. The resistance is due to the expression of an altered α -tubulin gene, the presence of which has been selected for as a direct result of repeated exposure of field populations of goosegrass to dinitroaniline herbicides. □

Methods

Maize transformation. DNA manipulations were performed by standard techniques¹⁵. All vector constructs (Fig. 3) were checked by restriction analysis and DNA sequencing. Maize BMS suspension culture cells were transformed using silicon-carbide-mediated DNA delivery as described¹⁶. Six independent transformation experiments were performed using each construct. Experimental controls performed for each transformation experiment were as follows: first, BMS cells plus BMS medium; second, BMS cells plus BMS medium plus $p\text{Bar}^+$ DNA ($p\text{Bar}^+$ is a construct consisting of the *bar* gene under the control of the 35S promoter inserted into plasmid pJR1); third, BMS cells plus BMS medium plus silicon carbide fibres on bialaphos; and fourth, BMS cells plus BMS medium on bialaphos. Genomic DNA was extracted from putative positives and analysed by PCR to check for *bar* gene insertion. Transgene copy number was determined by comparison of the signal on genomic Southern blots of the various clones probed with either *EiRtua1* or *EiStua1*, with gene copy number standards representing 1, 5 and 10 copies of the respective coding sequence.

Received 9 February; accepted 31 March 1998

- Morejohn, L. C., Bureau, T. E., Molé-Bajer, J., Bajer, A. S. & Fosket, D. E. Oryzalin, a dinitroaniline herbicide, binds to plant tubulin and inhibits microtubule polymerization *in vitro*. *Planta* **172**, 252–264 (1987).
- Parka, S. J. & Soper, O. F. The physiology and mode of action of the dinitroaniline herbicides. *Weed Sci.* **25**, 79–87 (1977).
- Appleby, A. P. & Valverde, B. E. Behaviour of dinitroaniline herbicides in plants. *Weed Technol.* **3**, 198–206 (1989).
- Mudge, L. C., Gossett, B. J. & Murphy, T. R. Resistance of goosegrass (*Eleusine indica*) to dinitroaniline herbicides. *Weed Sci.* **32**, 591–594 (1984).
- Vaughn, K. C., Marks, M. D. & Weeks, D. P. A dinitroaniline resistant mutant of *Eleusine indica* exhibits cross-resistance and supersensitivity to antimicrotubule herbicides and drugs. *Plant Physiol.* **83**, 956–964 (1987).
- Waldin, T. R., Ellis, J. R. & Hussey, P. J. Tubulin isotope analysis of two grass species resistant to dinitroaniline herbicides. *Planta* **188**, 258–264 (1992).

7. Akashi, T. *et al.* Effects of propyzamide on tobacco cell microtubules *in vivo* and *in vitro*. *Plant Cell Physiol.* **29**, 1053–1062 (1988).
8. Morejohn, L. C. & Fosket, D. E. The biochemistry of compounds with anti-microtubule activity in plant cells. *Pharmacol. Ther.* **51**, 217–230 (1991).
9. Villemur, R. *et al.* The α -tubulin gene family in maize (*Zea mays* L.): evidence for two ancient α -tubulin genes in plants. *J. Mol. Biol.* **227**, 81–86 (1992).
10. Hussey, P. J. *et al.* The β -tubulin gene family in *Zea mays*: two differentially expressed β -tubulin genes. *Plant Mol. Biol.* **15**, 957–972 (1990).
11. Jiang, C.-J. & Sonobe, S. Identification and preliminary characterisation of a 65 kDa higher-plant microtubule-associated protein. *J. Cell Sci.* **105**, 891–901 (1993).
12. Burns, R. G. & Surridge, C. D. in *Microtubules* (eds Hyams, J. S. & Lloyd, C. W.) 3–31 (Wiley-Liss, New York, 1994).
13. Nogales, E., Wolf, S. G. & Downing, K. H. Structure of the $\alpha\beta$ tubulin dimer by electron crystallography. *Nature* **391**, 199–203 (1998).
14. Lowe, J. & Amos, L. A. Crystal structure of the bacterial cell-division protein FtsZ. *Nature* **391**, 203–206 (1998).
15. Ausubel, F. M. *et al.* *Current Protocols in Molecular Biology* (Greene Publishing and Wiley InterScience, New York, 1992).
16. Frame, B. R. *et al.* Production of fertile transgenic maize plants by silicon carbide whisker-mediated transformation. *Plant J.* **6**, 941–948 (1994).
17. Last, D. I. *et al.* PEMU: an improved promoter for gene expression in cereal cells. *Theor. Appl. Genet.* **81**, 581–588 (1991).
18. Thompson, C. *et al.* Degradation of oxalic acid by transgenic oilseed rape plants expressing oxalate oxidase. *Euphytica* **85**, 169–172 (1995).
19. Field, J. *et al.* Purification of a RAS-responsive adenyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method. *Mol. Cell. Biol.* **8**, 2159–2165 (1988).

Acknowledgements. We thank J. Lowe and L. Amos for discussion about the structures of FtsZ and tubulin, E. Nogales and K. Downing for providing the coordinates of the $\alpha\beta$ -tubulin dimer and K. Roberts and H. Dickinson for critical reading of the paper. This work was supported by the Biotechnology and Biological Sciences Research Council (R.G.A., T.R.W., P.J.H.) and, in part, by Zeneca Agrochemicals (J.R., S.B.). This paper is dedicated to the memory of the late Ray Ellis with whom the work was initiated.

Correspondence and requests for materials should be addressed to P.J.H. (e-mail: p.hussey@rhnc.ac.uk). Genbank accession numbers are AJ005598 (*EiStua1*) and AJ005599 (*EiRtua1*).

An X-linked gene with a degenerate Y-linked homologue in a dioecious plant

David S. Guttman*† & Deborah Charlesworth*†

* Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637, USA

Most flowering plants are hermaphroditic, having flowers with both male and female parts. Less than 4% of plant species are dioecious (with individuals of separate sexes), and many of these species have chromosome-mediated sex determination. The taxonomic distribution of separate sexes and chromosomal sex-determination systems in the flowering plants indicates that plant sex chromosomes have evolved recently through replicated, independent events^{1–4}, contrasting with the ancient origins of mammalian and insect sex chromosomes. Plant sex chromosomes, therefore, offer opportunities to study the most interesting early stages of the evolution of sex chromosomes. Here we show that a gene encoding a male-specific protein is linked to the X chromosome in the dioecious plant *Silene latifolia*, and that it has a degenerate homologue in the non-pairing region of the Y chromosome. The Y-linked locus has degenerated as a result of nucleotide deletion and the accumulation of repetitive sequences. We have identified both the first X-linked gene and the first pair of homologous sex-linked loci to be found in plants. The homology between the active X-linked locus and the degenerate Y-linked locus supports a common ancestry for these two loci.

The widespread plant genus *Silene* (Caryophyllaceae) provides excellent opportunities for the study of sex-chromosome evolution. The several hundred *Silene* species are primarily gynodioecious (species having both female and hermaphroditic individuals) or hermaphroditic, but dioecy has evolved independently at least twice

Table 1 Loci analysed for sex linkage

Gene	Alternative name	Linkage*	Primer
<i>MROS1</i> †	<i>Men1</i>	A	+ ATG GCT CTC TCA TTC GCA ACT G – CGG CTG AAA CGT TGT CTC CAT C
<i>MROS2</i> †	<i>Men4</i>	A	+ GCA ACA TTC CGC CAC TCT CT – ACT CTG ATG TCG GAG CCA TAG C
<i>MROS3</i> †	<i>CCLS4, Men9</i>	X	+ GGA ACC CAA TCA CGC TTG A – TCA AGC ACG ACG AAC AAA CGC
<i>Sta1-18</i>	<i>Sta1-2, Sta1-12</i>	A	+ GGA AGC CGA TTT GGA GAA CAC G – CGC CAT CA CGA CCA AGT TAT G

* A, autosomal; X, X-linked. † Ref. 19. All loci were obtained from GenBank release 103.

within the genus⁵. The best studied plant sex-chromosome system is in the white campion, *S. latifolia* (also known as *Melandrium album*)^{4,6,7}. This diploid species is estimated to have diverged from its most recent, non-dioecious ancestor between 8 and 24 million years ago⁵, and has morphologically distinct X and Y sex chromosomes. Like most dioecious plants, females have the genotype XX, and males have the genotype XY. The *S. latifolia* Y chromosome differs from ancient mammalian Y chromosomes by being primarily euchromatic^{8,9}, but is similar in having a pseudo-autosomal region that pairs with the X chromosome⁴. It probably lacks essential genes, as androgenic haploid plants (having a Y chromosome but no X chromosome) are inviable¹⁰. At least three active loci are present on the Y chromosome^{4,11}: a female suppressor locus and loci responsible for anther maturation and early stamen development.

Theories for the evolution of sex chromosomes predict that the X and Y chromosomes evolved from a pair of homologous ancestors. Once sex-determining genes evolved, crossing over between the pair was reduced to prevent recombination between different sex determination loci^{12,13}. This, in turn, led to the slow evolutionary loss of function of the alleles on the chromosome found only in males (assuming male heterogamety), and finally to dosage compensation¹⁴. In plants, neither of these final two stages is known to have been reached, although there is evidence for heterochromatization of Y chromosomes in some *Rumex* species^{15,16}, and YY homozygotes are inviable in several dioecious plants^{4,10}.

We used multiplex single-strand-conformation polymorphism (SSCP)^{17,18} to identify genetic markers in several genes obtained from GenBank (Table 1), and to test for sex linkage of these genes in *S. latifolia*. Sex linkage was inferred from the transmission pattern of dominant DNA markers in a family array with six male and six female offspring. Bands present in the father, all female offspring, and none of the male offspring or the mother probably represent genes that are X-linked (Fig. 1). This band pattern could also occur for an autosomal marker if the father was heterozygous, but with a probability of $< 2 \times 10^{-4}$. The picture is more complicated if the mother is heterozygous for the marker, because bands will appear in both the mother and some male offspring. X-linked markers heterozygous in the mother can therefore be difficult to distinguish from autosomal markers heterozygous in both parents, or heterozygous in one and absent in the other. In the former case of X-linkage with a heterozygous mother only males that are recessive/hemizygous for a particular gene will lack the marker, whereas in the latter cases, both male and female double recessive offspring lacking the marker band can be found. Failure to see a sex-linked transmission pattern does not rule out sex linkage as the assay requires detectable polymorphism between the parents.

We observed a sex-linked transmission pattern only for *MROS3* (male reproductive organ specific gene)¹⁹ (Fig. 1). The bands indicated with arrows in Fig. 1c are present in the father, mother, all six female offspring, and four of six male offspring. We ruled out heterozygosity in both parents by cloning the parental *MROS3* polymerase chain reaction (PCR) products and sequencing multiple individual clones. Only one *MROS3* sequence was found out of seven paternal clones ($P = 0.0156$ of selecting the same sequence

† Present addresses: Department of Molecular Genetics and Cell Biology, University of Chicago, 1103 East 57th Street, Chicago, Illinois 60637, USA (D.S.G.); Institute for Cellular, Animal and Population Biology, University of Edinburgh, Ashworth Laboratory, King's Buildings, West Mains Road, Edinburgh EH9 3JN, UK (D.C.).

seven times, if two are present in equal proportions), whereas two sequences were found among the seven maternal clones. *MROS3* encodes a protein with no similarity to known proteins. Its expression is limited to the developing male flower tissue¹⁹.

We used anchored PCR (a modification of the Gibco BRL 5' rapid amplification of cloned ends (RACE) protocol requiring homology at only one priming site for successful amplification) to identify a male-specific homologue of the X-linked *MROS3*. Males and females yielded different amplification patterns after *Hae*III-digested genomic DNA was amplified with the *MROS3* minus primer (Fig. 2). A band of ~900 base pairs (bp) was consistently found in all males of the family, but none of the females. We isolated, cloned and sequenced this male-specific fragment (Fig. 3). A BLAST search indicated very strong homology between the 3' end of the *MROS3* database cDNA sequence and a 159-bp segment of the male-specific clone. In this region, 31 sites (19.5%) differed between

the X and Y sequences and there was a single insertion/deletion of three bases. There are 6.5 synonymous substitutions (out of 32.75 silent sites) and 24.5 non-synonymous substitutions (out of 120.25 replacement sites) in the male-specific clone, yielding a silent to substitution ratio of 0.974 ($K_a/K_s = 0.989$ with Jukes-Cantor correction²⁰). A ratio close to 1 is consistent with the protein's sequence evolving in a neutral manner²¹.

Outside the 159-bp region of high homology in *MROS3*-Y, there is a stretch of ~235 bp (corresponding to the central region in *MROS3*) that shows no homology with the X-chromosome locus. PCR amplification of *MROS3* from genomic DNA resulted in fragments consistent with the sizes predicted from the cDNA sequences¹⁹, indicating that this region in the Y-linked locus does not correspond to an intron in the X-linked *MROS3* gene. Primers designed to amplify only this region of *MROS3*-Y yielded identical bands in both males and females (data not shown), indicating that this region must be similar to an autosomal or X-linked locus. Database searches revealed no homology to any known nucleotide or protein sequences, and further characterization of the identity and origin of this region has not been possible. The rest of the *MROS3*-Y sequence (corresponding to the 5' end of *MROS3*-X) consists of sequential mononucleotide repeats averaging ten nucleotides in length.

Study of this first plant X-linked gene sequence shows that a degenerate Y-linked homologue exists for an X-linked locus. Pairs of active (X-linked) and degenerate (Y-linked) loci should exist in the non-pairing region of the sex chromosomes, because of the common autosomal ancestry of these two chromosomal segments^{13,14,22}, and such pairs have been found in human and mouse²³. Plant sex chromosomes have evolved recently, compared with those in animals (although some *Drosophila* species have recently evolved neo-sex chromosomes¹⁴), so Y-chromosomal loci in plants may be less likely to have lost expression. Furthermore, plant Y chromosomes may harbour genes that are actively expressed in the haploid male gametophytes during pollination²⁴⁻²⁶. Some Y-linked genes in plants will therefore be under selection pressure to preserve their function, even if they are in the differential segment of the Y chromosome. In the case of *MROS3*, expression is limited to the developing male flowers, so there is no reason to expect such selection on *MROS3*-Y. As a result, the gene indeed appears to have degenerated to a non-functional state. Homology between the active X-linked locus and the degenerate Y-linked locus is consistent with common ancestry for the X and Y chromosomes, although the ancestral source of

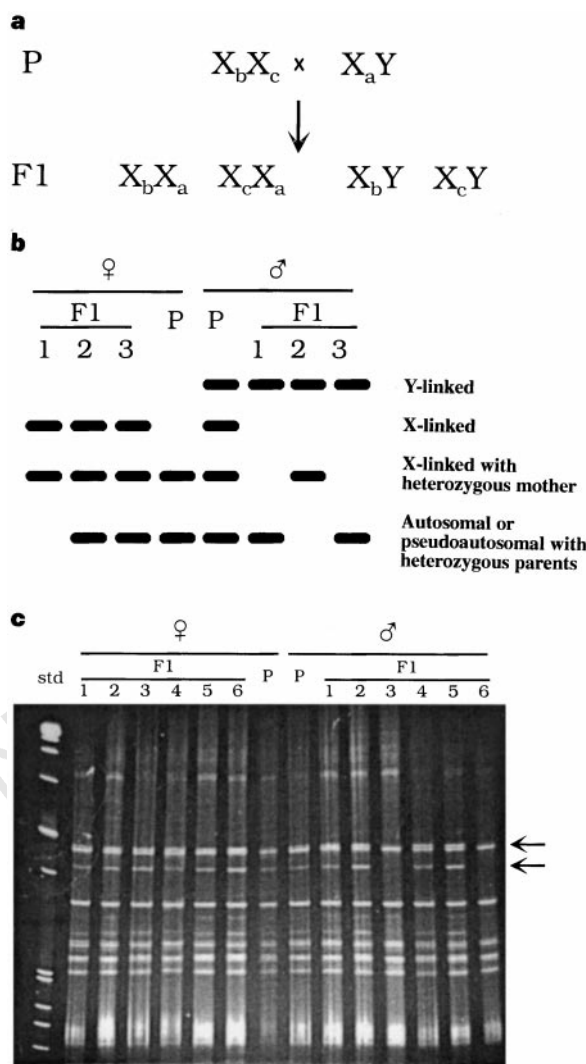


Figure 1 Sex linkage of *MROS3*. **a**, The segregation pattern of sex-linked loci used to identify sex linkage from SSCP data, where P is the male (XY) or female (XX) parent, and F₁ are the male and female offspring of the cross. **b**, An example of SSCP gel transmission patterns in different autosomal or sex-linked scenarios. The cartoon gel is ordered with three female offspring, the mother, father, and three male offspring. **c**, Multiplex cold SSCP gel photograph, obtained using *MROS3*. The lanes are ordered, from left to right, with a 1-kilobase ladder standard, six female offspring, mother, father, and six male offspring. The fragments that indicate X-chromosome linkage, by being absent in male offspring numbers 3 and 6, are marked with arrows.

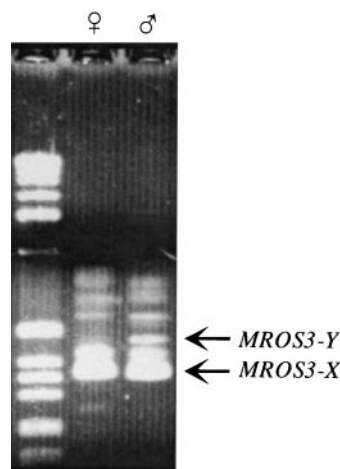


Figure 2 Gel photograph of the products of anchored PCR of *MROS3* from the mother and father of the family array used throughout the study. See Methods for details. One pattern was obtained from every male offspring, and a different one was seen in all the female offspring. The upper arrow indicates the male-specific, Y-linked locus. The lower arrow indicates the X-linked *MROS3*.

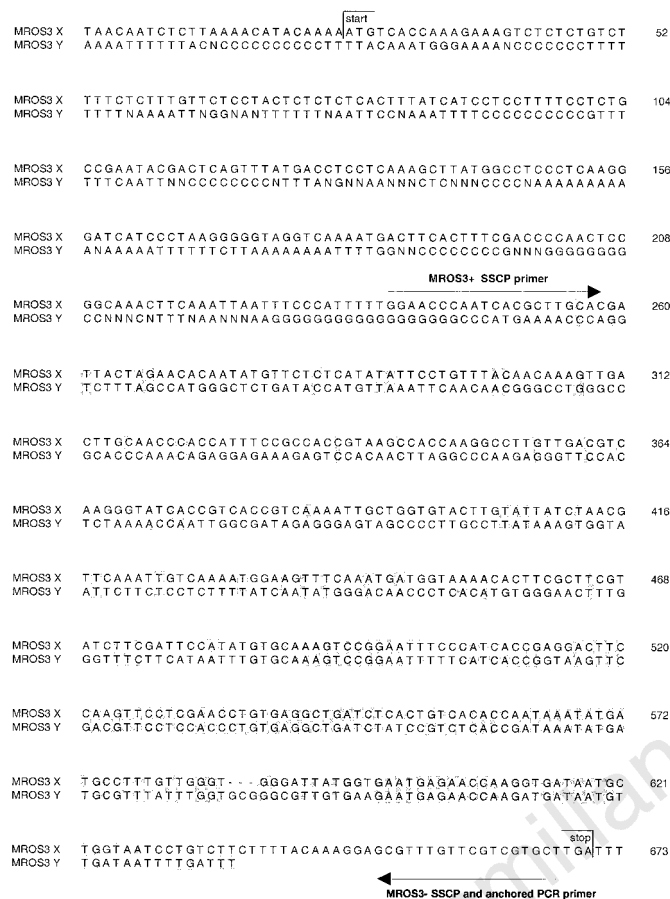


Figure 3 Sequence comparison of *MROS3-X* (upper sequence) and *MROS3-Y* (lower sequence). Matching sequences are shaded. A 3-bp deletion was included in the *MROS3-X* sequence after base 586 to preserve the alignment. The locations of the *MROS3+* and *MROS3-* primers used in the SSCP reaction are marked with arrows. The *MROS3-* primer was also used in the anchored PCR reaction. The first ~245 bases of *MROS3-Y* are a region of repetitive DNA, and, because of the difficulty in obtaining accurate data through these regions, the sequence shown is not meant to be taken as the exact sequence. The start and stop codons are denoted on the *MROS3-X* sequence.

the *MROS3-X* and *-Y* genes may be a chromosomal segment translocated onto an XY pair²⁶. To resolve this issue, studies of related species will be needed. □

Methods

Determination of sex linkage by multiplex cold SSCP. Five loci were tested for sex linkage (Table 1). PCR primers were designed from GenBank sequences to amplify regions of 200–400 bp in length. Genomic DNA was extracted from leaf tissue with a Puregene kit (Gentra Systems). PCR was performed according to manufacturer's instructions (Boehringer), using a cycle profile of: 95 °C 2 min; 35 cycles of 94 °C 30 s, between 50 °C and 60 °C (depending on the primer) 30 s, 72 °C 30 s; and a final cycle of 72 °C 5 min. PCR products were purified with Qiaquick columns (Qiagen) and restricted to reduce the size of the fragment, and SSCP was performed as described¹⁷. Multiplex SSCP differs from standard SSCP in that the banding pattern observed is complex enough that only the presence or absence of bands can be reliably scored, as opposed to standard SSCP where changes in the migration of specific bands can be scored. As a result, multiplex SSCP markers are most conservatively scored as dominant, rather than codominant markers.

Verification of X-linkage required checking for heterozygosity in both parents. The maternal and paternal *MROS3* PCR product were cloned into *EcoRV*-cut pZERO 2.1 (Invitrogen) vector according to manufacturer's instructions. Colony PCR was carried out on seven of the resulting colonies

from each parental plant (as described above, but with M13-20 and M13 reverse primers). The PCR products were cleaned with Qiaquick columns and 5 µl were sequenced with the TaqFS cycle sequencing kit (Applied Biosystems) and the vector primers.

Isolation of Y-chromosomal homologue. Male and female genomic DNA was digested with either *HaeIII* or *AluI*, and a poly (C) tail was added to the 3' end of each fragment with terminal deoxynucleotidyl transferase (Life Technologies) according to the manufacturer's instructions. The product was Qiaquick-purified and PCR-amplified with one gene-specific primer (either *MROS3+* or *MROS3-*) and a primer that binds to the poly(C)tail (AAP, Life Technologies). We used a two-step nested PCR amplification, because of the degree of divergence between the X and Y loci. PCR was performed as described above except that the first round cycle profile was: 95 °C 2 min; 35 cycles of 94 °C 30 s, 50 °C 30 s, 72 °C 1 min; and a final cycle of 72 °C for 5 min. 0.5 µl of the first-round product was used for the second round of amplification. A modified gene-specific primer that was extended by nine bases at the 3' end was used to allow higher annealing temperatures. The cycle profile was: 95 °C 2 min; 30 cycles of 94 °C 30 s, 55 °C 30 s increasing by 0.4 °C per cycle, 72 °C 1 min; and a final cycle of 72 °C for 5 min. The product of these reactions should yield a common band in both sexes, the product of the active X-linked gene, and, if the Y locus is inactive or altered from the state of the X-linked gene, there should also be a male-specific band. Figure 2 shows a band common to both males and females, of the size corresponding to the *MROS3-X*, and a male-specific band. Background bands can also be seen in both male and female reactions. These were due to the low annealing temperature necessary for the first round of PCR. Higher annealing temperatures eliminated these bands, but at the same time eliminated the partially diverged male-specific band. This result shows that *MROS3-Y* is unique, rather than being a duplication of *MROS3* on the Y chromosome. The male-specific product was cloned and five *MROS3-Y* colonies were sequenced as described above.

Received 8 January; accepted 26 February 1998.

- Charlesworth, D. In *Essays in Honour of John Maynard Smith* (eds Greenwood, P. J. & Slatkin, M.) 237–268 (Cambridge Univ. Press, 1985).
- Darwin, C. R. *The Different forms of Flowers on Plants of the Same Species* (John Murray, London, 1877).
- Mayer, S. S. & Charlesworth, D. Genetic evidence for multiple origins of dioecy in the Hawaiian shrub *Wikstroemia* (Thymelaeaceae). *Evolution* **46**, 207–215 (1992).
- Westergaard, M. The mechanism of sex determination in dioecious flowering plants. *Adv. Genet.* **9**, 217–281 (1958).
- Desfeux, C., Maurice, S., Henry, J.-P., Lejeune, B. & Gouyon, P.-H. Evolution of reproductive systems in the genus *Silene*. *Proc. R. Soc. Lond. B* **263**, 409–414 (1996).
- van Nigtevecht, G. Genetic studies in dioecious *Melandrium*. II. Sex determination in *Melandrium album* and *Melandrium dioicum*. *Genetica* **37**, 307–344 (1966).
- van Nigtevecht, G. Genetic studies in dioecious *Melandrium*. I. Sex-linked and sex-influenced inheritance in *Melandrium album* and *Melandrium dioicum*. *Genetica* **37**, 281–306 (1966).
- Vyskot, B., Araya, A., Veuskens, J., Negru, I. & Mouras, A. DNA methylation of sex chromosomes in a dioecious plant, *Melandrium album*. *Mol. Gen. Genet.* **239**, 219–224 (1993).
- Grant, S. *et al.* Genetics of sex determination in flowering plants. *Dev. Gen.* **15**, 214–230 (1994).
- Ye, D. *et al.* Sex determination in the dioecious *Melandrium*. I. First lessons from androgenic haploids. *Sex. Plant Reprod.* **3**, 179–186 (1990).
- Donnison, I. S., Stokoy, J., Vyskot, B., Saedler, H. & Grant, S. R. Isolation of Y chromosome-specific sequences from *Silene latifolia* and mapping of male sex determining genes using representational difference analysis. *Genetics* **144**, 1893–1901 (1996).
- Charlesworth, B. & Charlesworth, D. A model for the evolution of dioecy and gynodioecy. *Am. Nat.* **112**, 975–997 (1978).
- Charlesworth, B. Model for evolution of Y chromosomes and dosage compensation. *Proc. Natl Acad. Sci. USA* **75**, 5618–5622 (1978).
- Charlesworth, B. The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol.* **6**, 149–162 (1996).
- Rejón, C. R., Jamilena, M., Ramos, M. G., Parker, J. S. & Rejón, M. R. Cytogenetic and molecular analysis of the multiple sex chromosome system of *Rumex acetosa*. *Heredity* **72**, 209–215 (1994).
- Zuk, J. Autoradiographic studies in *Rumex* with special reference to sex chromosomes. *Chromosomes Today* **2**, 183–188 (1969).
- Guttman, D. S. *et al.* Multiple infections of *Ixodes scapularis* ticks by *Borrelia burgdorferi* as revealed by single-strand conformation polymorphism analysis. *J. Clin. Microbiol.* **34**, 652–656 (1996).
- Hongyo, T., Buzard, G. S., Calvert, R. J. & Weghorst, C. M. 'Cold SSCP': a simple, rapid and non-radioactive method for optimized single-stranded conformation polymorphism analyses. *Nucleic Acids Res.* **21**, 3637–3642 (1993).
- Matsunaga, S. *et al.* Isolation and developmental expression of male reproductive organ-specific genes in a dioecious campion, *Melandrium album* (*Silene latifolia*). *Plant J.* **10**, 679–689 (1996).
- Jukes, T. & Cantor, C. in *Mammalian Protein Metabolism* (ed. Munro, H. N.) 21–132 (Academic, New York, 1969).
- Nei, M. *Molecular Evolutionary Genetics* 1–512 (Columbia Univ. Press, New York, 1987).
- Charlesworth, B. The evolution of sex chromosomes. *Science* **251**, 1030–1033 (1991).
- Marshall Graves, J. A. The origin and function of the mammalian Y chromosome and Y-borne genes—an evolving understanding. *BioEssays* **17**, 311–321 (1995).
- Pedersen, S., Simonsen, V. & Loeschcke, V. Overlap of gametophytic and sporophytic gene expression in barley. *Theor. Appl. Genet.* **75**, 200–206 (1987).
- Stinson, J. R. *et al.* Genes expressed in the male gametophyte of flowering plants and their isolation. *Plant Physiol.* **83**, 442–447 (1987).

26. Tanksley, S. D., Zamir, D. & Rick, C. M. Evidence for extensive overlap of sporophytic and gametophytic gene expression in *Lycopersicon esculentum*. *Science* 213, 453–455 (1981).

Acknowledgements. We thank B. Charlesworth for input, counsel and support, and J. Greenberg, T. Morton and J. Mach for assistance with the manuscript. This work was supported by a grant from the NIH (to D.C. and B. Charlesworth).

Correspondence and requests for materials should be addressed to D.S.G. (e-mail: dguttman@midway.uchicago.edu).

Evidence for striatal dopamine release during a video game

M. J. Koepp, R. N. Gunn, A. D. Lawrence, V. J. Cunningham, A. Dagher, T. Jones, D. J. Brooks, C. J. Bench & P. M. Grasby

MRC Cyclotron Unit, Hammersmith Hospital, DuCane Road, London W12 0NN, UK, and Division of Neuroscience and Psychological Medicine, Imperial College School of Medicine, St Dunstan's Road, London W6 8RP, UK

Dopaminergic neurotransmission may be involved in learning, reinforcement of behaviour, attention, and sensorimotor integration^{1,2}. Binding of the radioligand [¹¹C]-labelled raclopride to dopamine D₂ receptors is sensitive to levels of endogenous dopamine, which can be released by pharmacological challenge^{3–8}. Here we use [¹¹C]-labelled raclopride and positron emission tomography scans to provide evidence that endogenous dopamine is released in the human striatum during a goal-directed motor task, namely a video game. Binding of raclopride to dopamine receptors in the striatum was significantly reduced during the video game compared with baseline levels of binding, consistent with increased release and binding of dopamine to its receptors. The reduction in binding of raclopride in the striatum positively correlated with the performance level during the task and was greatest in the ventral striatum. These results show, to our knowledge for the first time, behavioural conditions under which dopamine is released in humans, and illustrate the ability of positron emission tomography to detect neurotransmitter fluxes *in vivo* during manipulations of behaviour.

We used [¹¹C]-labelled raclopride (RAC) to detect changes in levels of extracellular dopamine induced by a behavioural task. During the first 50 minutes of a [¹¹C]RAC–PET scan, eight male volunteers played a video game, which involved learning to navigate a tank for a monetary incentive. This task is comparable to tasks in animal studies in which dopamine is released during the anticipatory or appetitive phase of motivated behaviour, where dopamine is involved in learning which environmental stimuli or actions predict rewarding or aversive outcomes^{2,9–11}. During a second [¹¹C]RAC–PET scan, subjects looked at an empty screen. The scanning order was randomized for each subject. Differences in [¹¹C]RAC-binding potential between scans were used to infer changes in levels of extracellular dopamine^{12,13}. Binding of [¹¹C]RAC to dopamine D₂

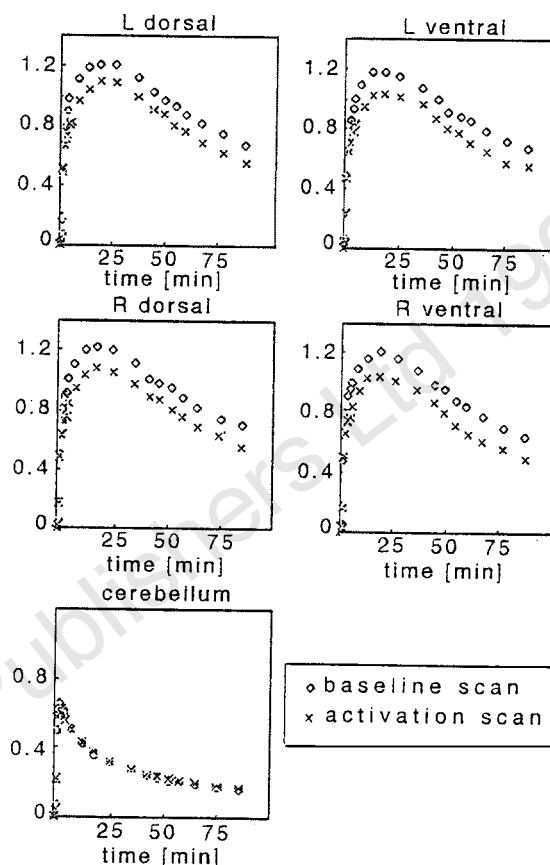


Figure 1 Mean time-activity curves for [¹¹C]RAC uptake, normalized for radioactivity injected, for the four striatal ROIs and the reference region (cerebellum). Data are given from time of radioligand injection to the end of scanning period (up to 90 min). R, right; L, left.

receptors was measured in the ventral and dorsal striata, which are areas involved in goal-directed motor behaviour^{2,14–16}.

Striatal [¹¹C]RAC-binding potential was reduced (analysis of variance (ANOVA) $F = 7.72$, $P < 0.01$) during the video game, particularly in the ventral striatum (Table 1). Our results are compatible with a task-related increase in levels of extracellular dopamine reducing the number of D₂-receptor sites available for binding to [¹¹C]RAC. The magnitude of change of [¹¹C]RAC-binding potential (ventral striatum mean, –13%; range, +8 to –42%) was considerably greater than the reported ‘within subject test/retest variation’ in striatal [¹¹C]RAC-binding potential (mean, 4–6%)^{17,18}, and was similar to that observed following intravenous injection of amphetamine⁸ (striatum mean, –16%; range, –3 to –24%) or methylphenidate⁶ (striatum mean, –23%; range, +3 to –46%). Microdialysis studies of non-human primates

Table 1 [¹¹C]RAC-binding potential, relative tracer delivery and size of the region of interest in striatal regions

	LD B	LD T	ΔLD (%)	RD B	RD T	ΔRD (%)	LV B	LV T	ΔLV (%)	RV B	RV T	ΔRV (%)
BP (s.d.)	2.47 (0.36)	2.23 (0.42)	–8.9 (16.4)	2.38 (0.34)	2.22 (0.41)	–6.1 (16.1)	2.22 (0.28)	1.93 (0.33)	–11.8 (18.8)	2.27 (0.31)	1.92 (0.35)	–13.9 (20.5)
R _i (s.d.)	0.98 (0.13)	0.88 (0.08)	–8.5 (13.8)	0.94 (0.12)	0.87 (0.07)	–6.5 (12.1)	0.98 (0.12)	0.89 (0.13)	–8.5 (14.4)	1.03 (0.11)	0.91 (0.12)	–10.5 (14.7)
ROI size (s.d.)	8,151 (711)	8,300 (846)	1.9 (8.2)	7,600 (567)	8,188 (574)	7.9 (7.2)	4,747 (628)	4,280 (612)	–8.7 (16.2)	4,692 (680)	4,215 (691)	–9.3 (15.7)

BP, [¹¹C]RAC-binding potential; R_i, relative tracer delivery; ROI, region of interest (mm³); s.d., standard deviation; L, left; R, right; D, dorsal; V, ventral; B, baseline conditions; T, task conditions. Changes in BP, R_i and ROI size between conditions (Δ) are given as a percentage change from the baseline, calculated as: $(T - B)/B \times 100$. R_i was significantly decreased during the video game ($F = 11.3$, $P = 0.001$), but reductions in R_i were not correlated with reductions in BP ($r^2 = 0.05$, $P = 0.24$). There was no difference in striatal ROI size across conditions ($P = 0.64$) and no correlation between changes of BP and ROI size ($r^2 = 0.02$, $P = 0.45$), indicating that head movement probably did not contribute significantly to our results.

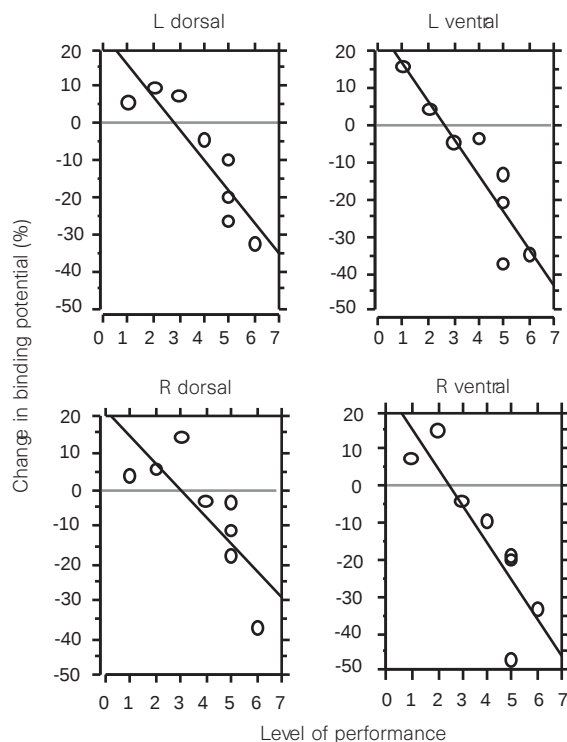


Figure 2 Percentage change in [^{11}C]RAC-binding potential between task and baseline conditions, plotted against performance level. A significant inverse correlation is seen in all striatal regions (Spearman rank correlation coefficients for left and right ventral and left dorsal striatum: $r = -0.86$, $P = 0.017$; right dorsal striatum: $r = -0.83$, $P = 0.020$).

indicate that a 1% decrease in striatal [^{11}C]RAC binding reflects at least an 8% increase in extracellular endogenous dopamine levels⁸. Thus, the 13% reduction in [^{11}C]RAC-binding potential in the ventral striatum reported here suggests at least a twofold increase in levels of extracellular dopamine. Computer simulations have shown that this magnitude of change should be detectable with [^{11}C]RAC-PET¹³.

After 50 min, the game ended, but the time-activity curves (TACs) for [^{11}C]RAC binding remained below the baseline curves without convergence (Fig. 1). A similar effect has been reported following pharmacological challenges^{4,19}, and may simply reflect the kinetic properties of [^{11}C]RAC and the diffusion and re-uptake kinetics of dopamine. Sustained alterations in dopamine concentrations after a period of behavioural manipulation have been described in the rat²⁰, providing a biological explanation for the continued separation of the TACs for [^{11}C]RAC binding after the video game ended.

There was a significant correlation between performance level achieved and reduced [^{11}C]RAC-binding potential in all striatal regions (Fig. 2); the significance of this correlation was confirmed by an independent analysis using statistical parametric mapping (SPM)²¹. SPM revealed that this significant correlation mainly encompassed the ventral striatum, predominantly the left side (Fig. 3). These results further validate the putative link between the behavioural manipulation and dopamine release, and complement electrophysiological studies of behaviour in awake animals, in which dopaminergic neurotransmission was associated with sensorimotor functions related to rewarding, aversive and stressful stimuli^{1,22,23}. In monkeys, most dopaminergic neurons in the ventral tegmental area and pars compacta are activated by unexpected primary appetitive rewards and reward-predicting cues^{1,9,15}.

Here, regional differences in [^{11}C]RAC displacement within the

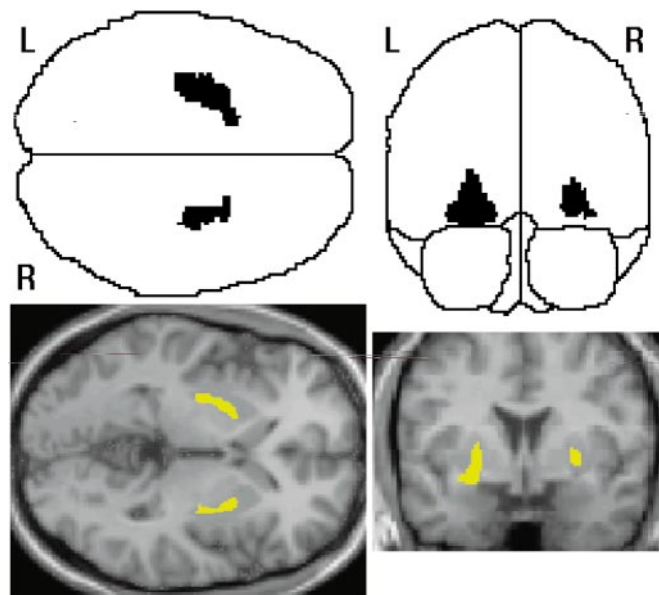


Figure 3 Regions of the brain in which there was a statistically significant correlation between reduced [^{11}C]RAC-BP and task performance; such a correlation was more pronounced in the ventral striatum. Upper row, the transverse and coronal glass brain views show those voxels with a significant inverse correlation of [^{11}C]RAC-BP with the highest performance level reached (threshold for display, $P < 0.05$). Lower row, three-dimensional SPM projections superimposed on representative transaxial and coronal magnetic resonance image brain slices (threshold for display, $P < 0.05$).

striatum might correlate with the role of dopamine in the dorsal and ventral striata². The dorsal striatum receives inputs from motor, sensory, premotor, and dorsal prefrontal cortices^{14,16}, whereas the ventral striatum receives afferent inputs from orbitofrontal cortex, amygdala, hippocampus, and anterior cingulate^{14,16}. On the basis of these anatomical connections, we interpret changes in ventral striatal [^{11}C]RAC binding to be related to affective components of the task, whereas dorsal striatal dopamine release may be related to sensorimotor coordination and response selection². This new method of detecting neurotransmitter release during behavioural manipulation extends the success of brain-perfusion mapping in humans to the study of a true 'cognitive neurochemistry of behaviour'. □

Methods

PET-scan acquisition. Eight healthy, male, right-handed volunteers (range 36–46 years of age) took part in the study (approved by the local Ethics Committee). Informed consent was obtained for all subjects. Each received two [^{11}C]RAC-PET scans (total injected dose of 16–20 mCi), one during the behavioural task (video game) and one under baseline conditions (blank screen). Subjects played the video game from 10 min before to 50 min after [^{11}C]RAC injection. PET scans were acquired on separate days using a 953B-Siemens/CTI PET camera in three-dimensional mode. Head movement during scanning was minimized by the use of a moulded head rest and external head markings.

Behavioural task. The video game involved moving a 'tank' through a 'battlefield' on a screen using a mouse with the right hand. Subjects had to collect 'flags' with the tank while destroying 'enemy tanks'. Enemy tanks could destroy the three 'lives' of the subjects' tank. If subjects collected all flags, they progressed to the next game level, which required more flags to be collected. A reward of £7 was given per game level achieved.

Region-of-interest (ROI) analysis. TACs of [^{11}C]RAC binding were derived for ventral and dorsal striata and cerebellum. From these TACs, binding potential (BP), and the relative rate of ligand delivery (R_l) in the striatum

compared to the cerebellum were estimated using a simplified reference region model^{24,25}. The model derives BP from the ratio of the volumes of distribution of the ligand in the striatum relative to the cerebellum. BP is a composite function of parameters, as follows:

$$BP = \frac{f_2 B_{\max}}{K_{D_{\text{Tracer}}} \left(1 + \sum_i \frac{F_i}{K_{D_i}} \right)}$$

where $B_{\max}$ is the total concentration of specific binding sites, $K_{D_{\text{Tracer}}}$ the equilibrium dissociation constant of the ligand, f_2 is the 'free fraction' of unbound ligand in the tissue, and F_i and K_{D_i} are the concentrations and equilibrium dissociation constants, respectively, of i competing endogenous ligands. Changes in BP are attributed to changes in F_i for endogenous dopamine. Striatal ROIs were outlined on an add-image of summated time frames, using an edge-fitting algorithm set at a fixed threshold (40%) of the image maximum. The ventral (comprising the ventral half of the putamen) and dorsal (comprising the dorsal half of the putamen and the body of the caudate nucleus) striata were operationally defined. The cerebellum was defined by cluster analysis²⁶. BP and R_1 values were calculated for the striatal ROIs using the TACs for [¹¹C]RAC binding up to 50 min after injection²⁵. Differences in [¹¹C]RAC-BP at baseline and during the task were tested with repeated-measure ANOVA, with three 'within-subject' factors (task versus baseline, left versus right hemisphere and dorsal versus ventral striatum). Spearman rank correlation coefficients were calculated for the relationship between changes in [¹¹C]RAC-BP and the highest performance level during the game for each ROI. **SPM analysis.** Parametric images of [¹¹C]RAC-BP²⁴ were analysed using SPM96 (ref. 21). The [¹¹C]RAC- R_1 images were used to define the stereotactic transformation parameters for the [¹¹C]RAC-BP images. Contrasts of the condition effects at each voxel of the [¹¹C]RAC-BP images were assessed using the t -value, with the highest performance level entered as a covariate of interest, giving a statistical image for each contrast.

Received 23 September 1997; accepted 20 March 1998.

- Schultz, W., Apicella, P. & Ljungberg, T. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J. Neurosci.* **13**, 900–913 (1993).
- Robbins, T. W. & Everitt, B. J. Functions of dopamine in the dorsal and ventral striatum. *Semin. Neurosci.* **4**, 119–127 (1992).
- Hume, S. P. *et al.* Quantitation of carbon-11 labelled raclopride in rat striatum using positron emission tomography. *Synapse* **12**, 47–54 (1992).
- Laruelle, M. *et al.* Microdialysis and SPECT measurements of amphetamine-induced dopamine release in non human primates. *Synapse* **25**, 1–14 (1997).
- Farde, L. *et al.* Positron emission tomography analysis of central D₁ and D₂ dopamine receptor occupancy in patients treated with classical neuroleptics and clozapine. *Arch. Gen. Psychiatry* **49**, 538–544 (1992).
- Volkow, N. D. *et al.* Imaging endogenous dopamine competition with [¹¹C]raclopride in the human brain. *Synapse* **16**, 255–262 (1994).
- Dewey, S. L. *et al.* Effects of central cholinergic blockade on striatal dopamine release measured with positron emission tomography in normal human subjects. *Proc. Natl Acad. Sci. USA* **90**, 11816–11820 (1993).
- Breier, A. *et al.* Schizophrenia is associated with elevated amphetamine induced synaptic dopamine concentrations: evidence from a novel positron emission tomography method. *Proc. Natl Acad. Sci. USA* **94**, 2569–2574 (1997).
- Ljungberg, T. J., Apicella, P. & Schultz, W. Responses of monkey dopamine neurons during learning of behavioural reactions. *J. Neurophysiol.* **67**, 145–163 (1992).
- Salamone, J. D., Cousins, M. S., McCullough, L. D., Carrier, O. D. L. & Berkovitz, R. J. Nucleus accumbens dopamine release increases during instrumental level pressing for food but not free food consumption. *Pharmacol. Biochem. Behav.* **49**, 651–660 (1994).
- Richardson, N. R. & Gratton, A. Behaviour-relevant changes in nucleus accumbens dopamine transmission elicited by food reinforcement: an electrochemical study in rat. *J. Neurosci.* **16**, 8160–8169 (1996).
- Fisher, R. E., Morris, E. D., Alpert, N. M. & Fischman, A. J. *In-vivo* imaging of neuromodulatory synaptic transmission using PET: a review of relevant neurophysiology. *Hum. Brain Mapp.* **3**, 24–34 (1995).
- Morris, E. D., Fisher, R. E., Alpert, N. M., Rauch, S. L. & Fischman, A. J. *In vivo* imaging of neuromodulation using positron emission tomography: optimal ligand characteristics and task length for detection of activation. *Hum. Brain Mapp.* **3**, 35–55 (1995).
- Graybiel, A. M., Aosaki, T., Flaherty, A. W. & Kimura, M. The basal ganglia and adaptive motor control. *Science* **265**, 1826–1831 (1994).
- Schultz, W. Dopamine neurons and their role in reward mechanisms. *Curr. Opin. Neurobiol.* **7**, 191–197 (1997).
- Brooks, D. J. The role of the basal ganglia in motor control: contributions from PET. *J. Neurol. Sci.* **128**, 1–13 (1995).
- Volkow, N. D. *et al.* Reproducibility of repeated measures of [¹¹C]raclopride binding in the human brain. *J. Nucl. Med.* **34**, 609–613 (1993).
- Logan, J. *et al.* Effects of blood flow on [¹¹C]raclopride binding in the brain: model simulations and kinetic analysis of PET data. *J. Cereb. Blood Flow Metab.* **14**, 995–1010 (1994).
- Endres, C. J. *et al.* Kinetic modelling of [¹¹C]raclopride: combined PET-microdialysis studies. *J. Cereb. Blood Flow Metab.* **9**, 932–942 (1997).
- Yamamoto, Y., Hori, K., Iwano, H. & Nomura, M. The relationship between learning-performance and dopamine in the prefrontal cortex of the rat. *Neurosci. Lett.* **177**, 83–86 (1993).

- Friston, K. J. *et al.* Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* **2**, 189–210 (1995).
- Thierry, A. M., Tassin, J. P., Blanc, G. & Glowinski, J. Selective activation of the mesocortical dopaminergic system by stress. *Nature* **263**, 242–244 (1976).
- Wise, R. A. The dopamine synapse and the notion of "pleasure centres" in the brain. *Trends Neurosci.* **3**, 91–94 (1980).
- Gunn, R. N., Lammertsma, A. A., Hume, S. P. & Cunningham, V. J. Parametric imaging of ligand-receptor binding in PET using a simplified reference region model. *Neuroimage* **6**, 279–287 (1997).
- Lammertsma, A. A. & Hume, S. P. Simplified reference tissue model for PET receptor studies. *Neuroimage* **4**, 153–158 (1996).
- Ashburner, J., Haslam, J., Taylor, C., Cunningham, V. J. & Jones, T. In *Quantification of Brain Function Using PET* (eds Myers, R., Cunningham, V., Bailey, D. & Jones, T.) 301–306 (Academic, London, 1996).

Acknowledgements. M.J.K. was supported by a grant from the Theodore and Vada Stanley Foundation Research Program; R.N.G., V.J.C., D.J.B. and P.M.G. were supported by the Medical Research Council; and A.D.L. was supported by a fellowship from the British Brain and Spine Foundation. We thank P. Dayan and L. Farde for discussions and comments on the manuscript; and K. Friston, A. Holmes and J. Ashburner for statistical advice and help with the SPM analysis.

Correspondence and requests for materials should be addressed to P.M.G. (e-mail: grasby@cu.rpm.ac.uk).

The role of dendrites in auditory coincidence detection

Hagai Agmon-Snir*, Catherine E. Carr† & John Rinzel*‡

* Mathematical Research Branch, NIDDK, National Institutes of Health, Bethesda, Maryland 20892, USA

† Department of Zoology, University of Maryland, College Park, Maryland 20742, USA

Coincidence-detector neurons in the auditory brainstem of mammals and birds use interaural time differences to localize sounds^{1,2}. Each neuron receives many narrow-band inputs from both ears and compares the time of arrival of the inputs with an accuracy of 10–100 μ s (refs 3–6). Neurons that receive low-frequency auditory inputs (up to about 2 kHz) have bipolar dendrites, and each dendrite receives inputs from only one ear^{7,8}. Using a simple model that mimics the essence of the known electrophysiology and geometry of these cells, we show here that dendrites improve the coincidence-detection properties of the cells. The biophysical mechanism for this improvement is based on the nonlinear summation of excitatory inputs in each of the dendrites and the use of each dendrite as a current sink for inputs to the other dendrite. This is a rare case in which the contribution of dendrites to the known computation of a neuron may be understood. Our results show that, in these neurons, the cell morphology and the spatial distribution of the inputs enrich the computational power of these neurons beyond that expected from 'point neurons' (model neurons lacking dendrites).

Over the past 40 years it has become widely accepted that dendrites play a major role in neuronal computation⁹. Despite intensive efforts to decipher this role^{10–16}, however, the contribution of the dendrites to the function of the single neuron remains elusive. Nevertheless, the existence of different dendritic geometries and their plausible effect on computation have been used as evidence for dendritic computation^{11,12,17}. As analysis of dendritic computation is most powerful when the role of the neuron is understood, we used brainstem auditory coincidence detectors to demonstrate the computational advantages of having synaptic inputs on the dendrites rather than on the cell body.

Coincidence detectors of the auditory brainstem are binaural neurons that respond maximally when they receive simultaneous inputs from the two ears. This condition is met when delay line inputs from each ear exactly compensate for a delay introduced by an interaural time difference (ITD), the time difference between the

‡ Present address: New York University, Center for Neural Science and Courant Institute of Mathematical Sciences, New York 10003, USA.

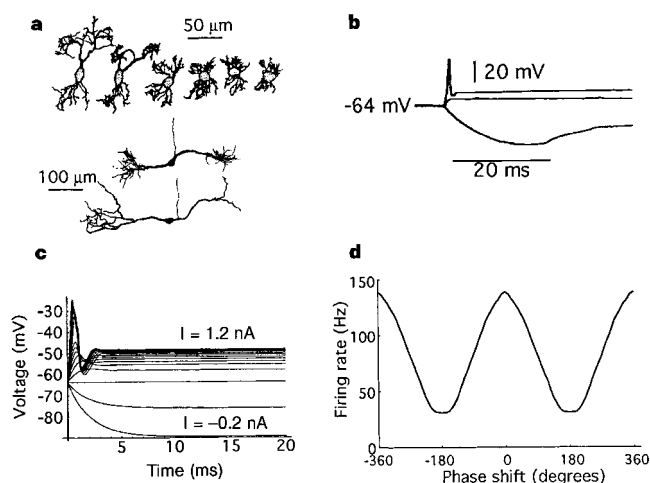


Figure 1 A point-neuron model for auditory coincidence detectors. **a**, Avian (top) and mammalian (bottom) low-frequency coincidence detector neurons. The stimulus frequency of the chicken nucleus laminaris cells increases from left to right (adapted from refs 7, 30). The bipolar architecture and the segregation of the inputs from both ears is common to both mammalian and avian coincidence detectors. **b**, Superimposed voltage responses during application of subthreshold and suprathreshold current steps to nucleus laminaris neurons²⁰. **c**, Responses of the point-neuron model to step current injections (see Methods). The output of the coincidence detector is phase-locked to their inputs (with some jitter; Fig. 2b). **d**, ITD curve in response to 1 kHz stimulus frequency shows that the point-neuron model performs like coincidence detectors in mammals and birds^{3,4,6,20}.

two ears)^{1–4}. These coincidence-detector neurons are particularly suitable for an analysis of dendritic function. First, they have stereotyped bipolar dendrites (Fig. 1a)^{7,8}. Second, the inputs from each ear are segregated on the dendrites, with inputs from the ipsilateral ear terminating on the dorsal dendrites and inputs from the contralateral ear on the ventral dendrites^{7,8}. Third, the length of the dendrites increases with decreasing best frequency of the sound stimulus in the chick^{7,18} (Fig. 1a), suggesting a computational role for the dendrites. Fourth, the electrophysiological properties of these neurons have been well characterized^{19,20}. As is appropriate for timing devices, these coincidence detectors do not respond to slowly varying inputs and fire one or few spikes in response to a step current injection (Fig. 1c).

We used data from the neurons of the ITD circuit to construct a minimal biophysical model that mimicked the essential properties of the coincidence detectors (Fig. 1b, c; see Methods)^{4,19–22}. This model, like earlier models^{23–26}, detected coincidences when inputs from both ears arrived directly at its soma (Fig. 1d). In the present model, however, we were able to investigate the contribution of dendrites. We modelled two simple dendrites as unbranched cables (electronic length, 0.2λ), on the basis of measured dendritic lengths and the biophysical properties of the coincidence detectors (see Methods)^{7,20}. The model cell received 12 inputs, 6 from each ear. In some cases, all 12 inputs were located on the soma, and in

other cases, all 6 inputs from the left ear were located on one dendrite, 0.1λ from the soma, and all 6 from the right ear on the other dendrite (Fig. 2a). The inputs were phase-locked (Fig. 2b; see Methods).

When a sound moves around the head, there is a phase shift between the two ears, and the response of the coincidence detector varies in cyclic fashion between best ITD (maximum) and worst ITD (minimum; Fig. 1d and 2c). Any mechanism that increases the maximum and at same time decreases the minimum of these ITD plots improves the detection of sound localization. Indeed, in model ITD plots, the contrast between the maximal and minimal spike rates was greater when the inputs from each ear were segregated and located on the dendrites (Fig. 2c). This improvement in coincidence detection produced by locating the inputs on the dendrites was robust⁹. We showed this by changing the single (unitary) synaptic conductance (USC) as a parameter for several different input configurations, as maximal and minimal spike rates also depended upon USC amplitude (Fig. 2d). For each configuration, maximum and minimum spike rates increased with increasing synaptic conductance. Best coincidence detection (Fig. 2d, dotted line) was achieved when the inputs were segregated on the dendrites, and when the dendrites were thin (see Methods).

This improvement in ITD coding is based on local computation at the dendrites. In general, synaptic inputs sum nonlinearly,

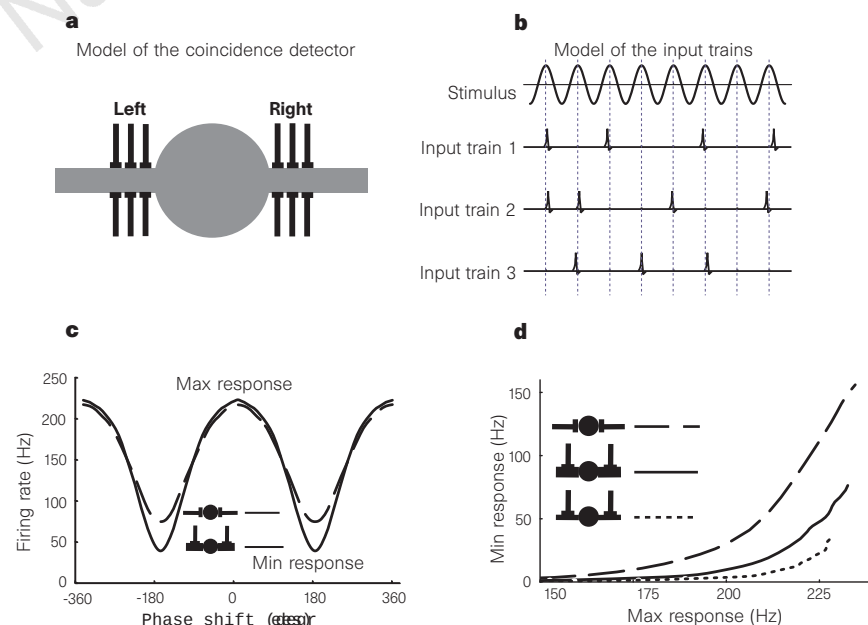


Figure 2 A bipolar-neuron model for coincidence detection. **a**, The bipolar-neuron model consists of a soma (centre) and two cylindrical dendrites (see Methods). **b**, Inputs resembled *in vivo* inputs (see Methods). **c**, Model ITD curves for a 500-Hz stimulus showed an improvement in ITD coding when dendrites were added (0.2λ thick ($4 \mu\text{m}$) passive dendrites). The contrast between the maximum spike rate (0° delay) and the minimum spike rate (180° delay) was larger when the inputs were segregated on the dendrites (solid line; USC = $0.022 \mu\text{S}$), compared with the point-neuron model (dashed line; USC = $0.01 \mu\text{S}$). **d**, Parametric plot of different input configurations shows the relationship between maximum and minimum spike rates as the USC changes. Dashed line, inputs on soma, USC range 0.007 – $0.011 \mu\text{S}$; solid line inputs on thick ($4 \mu\text{m}$ diameter) dendrites, USC range 0.013 – $0.024 \mu\text{S}$; dotted line, inputs on thin ($2 \mu\text{m}$ diameter) dendrites, USC range 0.014 – $0.026 \mu\text{S}$. Dendritic USCs were made larger than somatic USCs to provide similar maximum responses.

because the driving force for excitatory synaptic currents decreases with depolarization. Hence, the net synaptic current from several inputs arriving simultaneously at nearby sites on the same dendrite is smaller than the current generated if these inputs arrive at different dendrites⁹. Figure 3a illustrates how the segregation helps to improve coincidence detection. Thus, the conductance threshold, or minimum synaptic conductance needed to trigger a somatic action potential, is higher when the synaptic events are on the same dendrite, compared with when they are split between the bipolar dendrites. In the bipolar model, when the inputs are segregated and the phase shift is 0°, inputs arrive simultaneously at both dendrites and it is comparably easy to generate an action potential. When the phase shift is 180°, at every half cycle inputs arrive on only one dendrite and the conductance threshold is large. In the point neuron, however, the conductance threshold does not vary when the phase shifts are 0° and 180°. Thus, the difference between the firing rates in the 0° and the 180° phase shift cases is larger for the bipolar neuron with segregated inputs than for the point neuron.

Action-potential thresholds depended on whether the inputs were on the cell body or the dendrites (Fig. 3b). We used the concept of conductance threshold to explain the difference between segregated and non-segregated inputs. For each model configuration, points above the curve represent combinations of suprathreshold synaptic inputs from left (g_L) and right (g_R) ears. When binaural inputs arrive at the same location, action potentials fire when $g_L + g_R$ is larger than a fixed conductance threshold (Fig. 3b, straight dashed line). Conductance-threshold plots are not linear when the inputs from each ear are segregated on different dendrites. We used the biophysical model to calculate these thresholds and

show that the conductance threshold ($g_L + g_R$) for equally distributed inputs ($g_L = g_R$; points I and II in Fig. 3b) is smaller than the threshold (g_{Th}) for inputs arriving from only one side ($g_L = 0$ or $g_R = 0$; points III and IV). Thus, the model cell was most likely to fire action potentials when stimulated by coincident inputs from each ear.

As dendritic length varied with best frequency in the chicken⁷, we also investigated the effects of stimulus frequency and dendritic length on coincidence detection. The performance of the model (Fig. 2d) deteriorated with inputs of higher frequencies (Fig. 3c; the curves for higher frequencies were to the left of those of lower frequencies, reflecting poorer coincidence detection). Furthermore, at higher frequencies the contrast between the maximum and minimum spike rates did not change much when the inputs were on the dendrites. The main reason for the reduction in dendritic advantage at higher stimulus frequencies was input jitter. At higher frequencies, jitter resulted in an overlap in time between inputs from both ears, even when there was a 180° phase shift between them. The effect of this overlap on the conductance threshold was comparably large when the inputs were on the dendrites; even a small 'erroneous' input from one ear significantly lowered the conductance threshold for the inputs from the other ear (compare, for example, points III and V in Fig. 3b). As a result, at high frequencies the firing rate in the 180° mode could be even larger than in the somatic model, making the dendrites a burden instead of an advantage. This prediction of a negligible dendritic length required to detect higher frequencies conforms to the observation of short dendrites on high-frequency-coincidence detectors in chickens⁷.

From this analysis, we understood that in the presence of jitter,

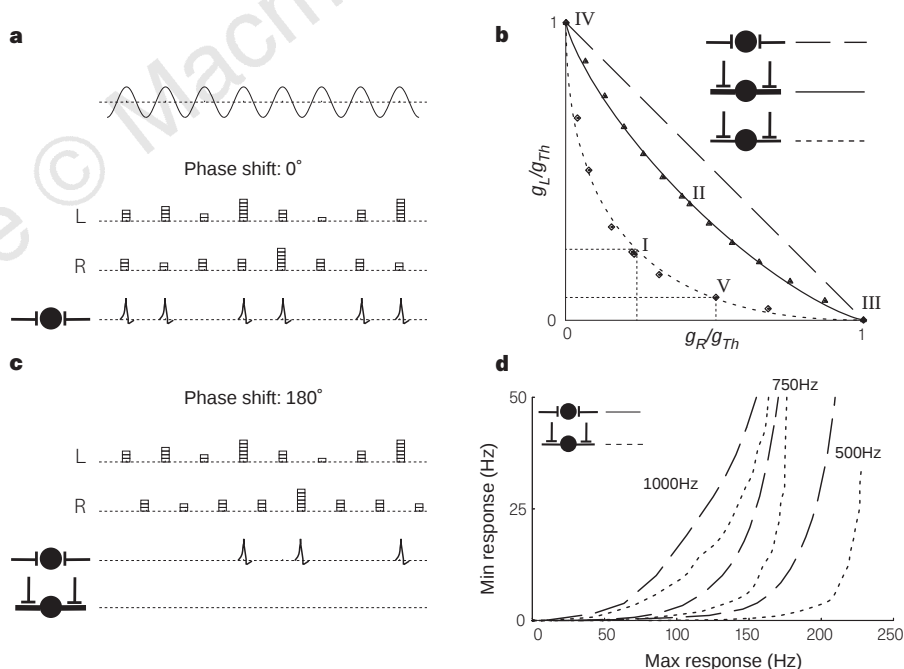


Figure 3 Biophysical mechanism and frequency dependence. **a**, The role of nonlinear integration of synaptic inputs in enhancement of coincidence detection is explained in a schematic diagram. The sinusoidal stimulus (top curve) leads to a binomially distributed random number of synaptic events from the left (L) and right (R) ears at every cycle (depicted as stacks of blocks). Suppose that six simultaneous synaptic events were just enough to trigger a postsynaptic spike in the point neuron. More firings must occur when the phase shift is 0° than when it is 180° (in the example, six spikes in the 0° case and three in the 180° case). In the bipolar neuron, the USC can be adjusted so that the firing rate for phase shift of 0° equals that of the point neuron. Hence, three inputs from the left arriving at the same time as three inputs from the right would fire the bipolar cell. Six inputs

arriving from only one side, however, would not fire the cell, and the 180° firing rate would be smaller. **b**, Schematic threshold lines for different input configurations. Using the biophysical model, we calculated thresholds for the cases of Fig. 2d (triangles, thick dendrites; diamonds, thin dendrites). g_{Th} is defined as the conductance threshold for inputs arriving from only one side. We fitted those points to the equation $g_L^\alpha + g_R^\alpha = g_{Th}^\alpha$, and good fit was obtained for $\alpha = 0.77$ and 0.48, respectively. The straight line therefore represents the case $\alpha = 1$. **c**, The effect of frequency on coincidence detection in the biophysical model. The minimum and maximum spike rates for the somatic and dendritic configurations are shown for three frequencies. The dendritic model is as in Fig. 2d with thin ($2\mu\text{m}$ diameter) dendrites.

severe nonlinear summation is a disadvantage. With long dendrites, larger voltages were required at the dendritic site to trigger a spike at the soma, as the opposite dendrite served as a current sink and the voltage gradient from the dendritic site to the soma was steeper. As a result, the summation of inputs at the dendritic site was more nonlinear than for shorter dendrites, and could worsen the coincidence detection at high stimulus frequencies. For example, when $f_{\text{stim}} = 500$ Hz, the best coincidence detection occurred when the dendrites were about 0.4λ each, whereas optimal dendritic lengths were shorter for higher frequencies.

A simple bipolar integrate-and-fire model that retained the biophysical model's essential features improved our understanding of how the nonlinear summation on bipolar dendrites enhances coincidence detection and why the optimal dendritic length was shorter for higher frequencies. Suppose that, within a given stimulus cycle, left-ear afferents evoked a net synaptic input conductance transient of size g_L , and right-ear afferents evoked g_R . In response, the neuron fired with probability $P(g_L, g_R)$. If the afferent inputs arrived at a single site and if the noise-free neuron had a sharp fixed threshold, then $P = 1$ if $g_L + g_R \geq g_{\text{Th}}$ and $P = 0$ otherwise, where g_{Th} is the conductance threshold at this input site. When the inputs were segregated on the bipolar dendrites, the nonlinear integration

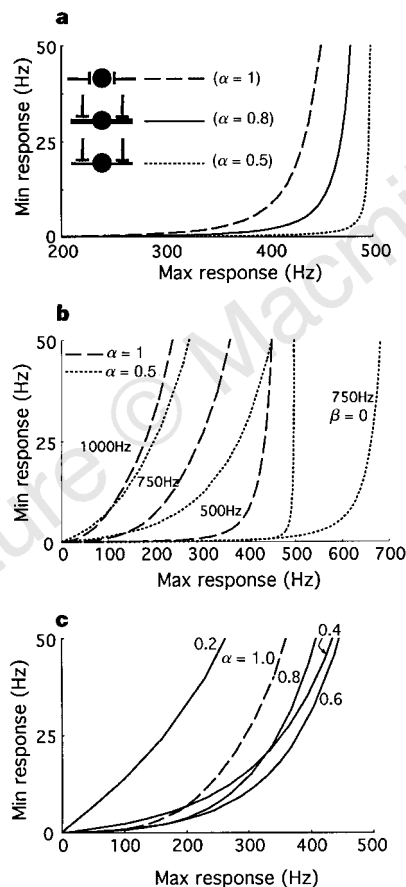


Figure 4 Bipolar integrate-and-fire model. **a**, Contrast enhancement in the integrate-and-fire model demonstrated with parametric plots for different input configurations (as in the biophysical model; Fig. 2d). We assumed 12 input lines, 6 from each side; $f_{\text{stim}} = 500$ Hz and $f_{\text{pre}} = 350$ Hz. **b**, The effect of jitter on coincidence detection for various frequencies is demonstrated in the modified integrate-and-fire model. $\beta = 0$ for 500 Hz, 0.15 for 750 Hz, and 0.3 for 1,000 Hz. The dendritic ($\alpha = 0.5$) case for 750 Hz with $\beta = 0$ is also shown. **c**, The effect of α on coincidence detection in the modified integrate-and-fire model. $f_{\text{stim}} = 750$ Hz, $\beta = 0.15$. The optimal coincidence-detection properties are found here for $\alpha = 0.6$. For $\alpha = 0.2$, coincidence detection is worse than for $\alpha = 1$, reflecting the loss of dendritic advantage.

of inputs seen in the biophysical model could be mimicked by using the condition $g_L^\alpha + g_R^\alpha \geq g_{\text{Th}}^\alpha$, where $\alpha \leq 1$ depended on the nonlinear integration at the dendrites (Fig. 3b). Assuming n afferents from each ear, and independence between synaptic events in different stimulus cycles and in different input trains, the probability, $b_{i,n}$, that i synaptic events were delivered by one ear's afferents in a stimulus cycle could be approximated using a binomial distribution. We calculated the probability for postsynaptic firing in a stimulus cycle when there is no phase shift between the left and right inputs (P_0) and when the binaural phase shift is 180° (P_{180}); these probabilities depend on g_{syn} , the USC (see Methods). The cell's mean firing rate is $P_0 f_{\text{stim}}$ for the 0° phase shift case and $P_{180} f_{\text{stim}}$ for the 180° case. This simple analytical model yields results in qualitative agreement with those from our minimal biophysical model (Fig. 4a). It embodies the combined essential effects due to randomness in the number of the synaptic events per cycle and the nonlinear integration caused by the spatial distribution of the afferents.

The role of jitter was demonstrated in the integrate-and-fire model by a simple modification. We assumed that in the 180° phase shift case, a fraction, β , of the input conductance from one ear might affect the integration of the inputs from the other ear (see Methods). Analysis of the input trains of the biophysical model showed that β is comparably small for $f_{\text{stim}} = 500$ Hz, whereas it can approach 0.4 when $f_{\text{stim}} = 1,000$ Hz. In this modified model, coincidence detection was worse at higher frequencies, and the dendritic advantage was diminished (Fig. 4b). Without considering jitter ($\beta = 0$ for all frequencies), putting inputs on dendrites ($\alpha < 1$) was advantageous even at high frequencies. Using the modified model it was easy to demonstrate that when β was non-zero, there was an optimal α for best coincidence detection (Fig. 4c). As longer dendrites have smaller values of α , there is an optimal dendritic length for each frequency, and this optimal length decreases with higher frequencies. These predictions match the changes in dendritic length observed in the chicken coincidence detectors⁷ (Fig. 1a).

We have shown how dendrites improve coincidence detection in the cells of the auditory brainstem, using basic biophysical mechanisms to explain their computational role. One mechanism is the segregation of the inputs on the dendrites, allowing nonlinear integration between the inputs from left and right ears. The second mechanism is a modulation of the nonlinearity of the integration by using the dendrites as current sinks for each other. The computational module isolated in the auditory coincidence detectors might be used in other neurons with branched dendritic trees, perhaps as part of a more complex computation. For example, in cortical neurons, the segregation of inputs on the two main branches of the pyramidal-cell apical tree might use a similar computational module, with the same biophysical mechanism as the brainstem auditory cells. The segregation of inputs on the many basal dendrites might also be involved in such a mechanism. The investigation of other cells with known functions could yield a set of plausible 'computational building blocks' that might help to decipher the dendritic code in more complex cells with unknown functions. □

Methods

Point-neuron model. Data from nucleus magnocellularis and nucleus laminaris cells were used to construct the model, which mimicked the properties of coincidence detectors, including phase-locking to about $1.5 \text{ kHz}^{19-22,27}$. The fast potassium currents ($\tau < 1 \text{ ms}$) that repolarize the cell were modelled, whereas slow potassium currents ($\tau > 10 \text{ ms}$) were substituted by a steady-state conductance. The model equations were Morris-Lecar type^{28,29}: $dV/dt = (-1/C_m)[\bar{g}_{\text{Na}} m_\infty(V)(V - V_{\text{Na}}) + g_{\text{syn}}(V - V_{\text{syn}}) + \bar{g}_{\text{K}} w(V - V_{\text{K}}) + g_L(V - V_L)]$; $dw/dt = \phi[(w_\infty(V) - w)/\tau_w(V)]$; where $C_m = 0.0147 \text{ nF}$, $\bar{g}_{\text{Na}} = 33 \text{ nS}$, $V_{\text{Na}} = 20 \text{ mV}$, $V_{\text{syn}} = 0 \text{ mV}$, $\bar{g}_{\text{K}} = 237 \text{ nS}$, $V_{\text{K}} = -70 \text{ mV}$, $g_L = 7 \text{ nS}$, $V_L = -62.5 \text{ mV}$, $\phi = 1.0$, $m_\infty(V) = 1/[1 + \exp\{(V - V_1)/V_2\}]^2$, $w_\infty(V) = 1/[1 + \exp\{(V - V_3)/V_4\}]^2$, $\tau_w(V) = 1/\cosh\{(V -$

$V_3/V_6\}$, and, in mV, $V_1 = -42.5$, $V_2 = -1$, $V_3 = -43.0$, $V_4 = -4$, $V_5 = -60.0$, and $V_6 = 64$. Simulation results were obtained by numerical integration of differential equations over 10 s.

Bipolar-neuron model. See Fig. 2a. The soma was the same as in the point-neuron model (20 μm diameter, $R_{\text{soma}} = 135 \text{ M}\Omega$); the dendritic diameter was either 4 μm (thick dendrites) or 2 μm (thin dendrites), with axial resistivity $R_i = 200 \text{ }\Omega \text{ cm}$ and membrane resistance $R_m = 1,700 \text{ }\Omega \text{ cm}^2$. For a dendrite of diameter 4 μm , $\lambda = 290 \text{ }\mu\text{m}$, whereas for a dendrite of diameter 2 μm , $\lambda = 200 \text{ }\mu\text{m}$. Parameters were based on recordings from chicken coincidence detectors²⁰. Dendrites were modelled by 0.05- λ -connected compartments and had either active membrane (identical to the point neuron) or passive membrane (voltage-dependent conductances fixed to their resting values). Because synaptic inputs arrived at the cell in every cycle, it was insufficient to use a simple threshold function to recognize action potentials in the somatic response. We therefore added a long axon with a higher density of voltage-dependent conductances. We counted only action potentials that propagated.

Synaptic-input model. See Fig. 2b. For every input train, at every stimulus cycle, the probability of an input arriving was defined as $f_{\text{pre}}/f_{\text{stim}}$, where f_{stim} was the stimulus frequency and $f_{\text{pre}} (\leq f_{\text{stim}})$ was the average spike rate of the input train. $f_{\text{pre}} = 350 \text{ Hz}$ for all stimulus frequencies²⁷. The stimulus cycles were regulated as independent events. To account for the jitter in the phase-locking of the inputs to the stimulus, measured by vector strength (VS)³, we shifted each input in time from the beginning of the cycle by a random variable $t_{\text{shift}} \sim N(m = 0, \sigma = (1,000/f_{\text{stim}})\{\sqrt{1 - 2\ln(VS)/(2\pi)}\} \text{ ms})$. For $VS \geq 0.2$, this resulted in input trains with the required VS. Except in Fig. 1d, we used $VS = 0.7$. In Fig. 1d, $VS = 0.8$. Synaptic inputs were rectangular conductance changes, 0.4 ms wide.

Bipolar integrate-and-fire model. The probability that the coincidence-detector neuron would fire in a given stimulus cycle was assumed to be $P(g_L, g_R) = 1/(1 + \exp\{[1 - (g_L/g_{\text{Th}})^\alpha - (g_R/g_{\text{Th}})^\alpha]/k\})$, where g_L and g_R were the total synaptic input conductance during this cycle from the left and right ears, respectively. We used $k = 0.05$ and $g_{\text{Th}} = 1$ (the positive parameter k determines the steepness of the sigmoid threshold function). By using a sigmoid $P(g_L, g_R)$ (rather than the condition $g_L^\alpha + g_R^\alpha \geq g_{\text{Th}}^\alpha$, which is equivalent to the sigmoidal function when k approaches 0), we approximately accounted for the effects of small intrinsic noise, jitter in a composite input, and a graded threshold for the spike-generating mechanism. The probability $b_{i,n}$ (see text) was calculated by $b_{i,n} = \binom{n}{i} (f_{\text{pre}}/f_{\text{stim}})^i (1 - f_{\text{pre}}/f_{\text{stim}})^{n-i}$. Then, P_0 is

$$P_0 = \sum_{i=0}^n \sum_{j=0}^n b_{i,n} b_{j,n} P(i g_{\text{syn}}, j g_{\text{syn}})$$

where input combinations are summed with index i for the left side and j for the right. On the other hand, if the binaural phase shift was 180° , the probability to fire during one cycle (that is, the probability that inputs from the left or the right ear will cause firing) was approximated by

$$P_{180} = 2 \sum_{i=0}^n b_{i,n} P(i g_{\text{syn}}, 0)$$

provided that P_{180} was small enough. In the modified version of the model, the jitter was accounted for by using instead

$$P_{180} = 2 \sum_{i=0}^n \sum_{j=0}^n b_{i,n} b_{j,n} P(i g_{\text{syn}}, j g_{\text{syn}})$$

Received 7 January; accepted 16 March 1998.

- Jeffress, L. A. A place theory of sound localization. *J. Comp. Physiol. Psychol.* **41**, 35–39 (1948).
- Carr, C. E. Timing mechanisms in the CNS. *Annu. Rev. Neurosci.* **16**, 223–243 (1993).
- Goldberg, J. M. & Brown, P. B. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J. Neurophysiol.* **32**, 613–636 (1969).
- Carr, C. E. & Konishi, M. A circuit for detection of interaural time differences in the brainstem of the barn owl. *J. Neurosci.* **10**, 3227–246 (1990).
- Overholt, E. M., Rubel, E. W. & Hyson, R. L. A delay-line circuit for coding interaural time differences in the chick brain stem. *J. Neurosci.* **12**, 1698–1708 (1992).
- Yin, T. C. T. & Chan, J. C. K. Interaural time sensitivity in the medial superior olive of the cat. *J. Neurophysiol.* **64**, 465–488 (1990).
- Smith, Z. D. J. & Rubel, E. W. Organization and development of brain stem auditory nuclei of the chicken: dendritic gradients in nucleus laminaris. *J. Comp. Neurol.* **186**, 213–239 (1979).
- Stotter, W. A. An experimental study of the cells and connections of the superior olivary complex of the cat. *J. Comp. Neurol.* **98**, 401–432 (1953).
- Segev, I., Rinzel, J. & Shepherd, G. *The Theoretical Foundation of Dendritic Function: Selected Papers of Wilfrid Rall with Commentaries* (MIT Press, Cambridge, MA, 1995).

- Mel, B. Information processing in dendritic trees. *Neural Comp.* **6**, 1031–1085 (1994).
- Segev, I. In *The Handbook of Brain Theory and Neural Networks* (ed. Arbib, M.) 282–289 (MIT Press, Cambridge, MA, 1995).
- Pinsky, P. F. & Rinzel, J. Intrinsic and network rhythmogenesis in a reduced Traub model for CA3 neurons. *J. Comput. Neurosci.* **1**, 39–40 (1994).
- Mainen, Z. F. & Sejnowski, T. J. Influence of dendritic structure on firing pattern in model neocortical neurons. *Nature* **382**, 363–366 (1996).
- Softky, W. R. & Koch, C. The highly irregular firing of cortical cells in inconsistent with temporal integration of random EPSPs. *J. Neurosci.* **13**, 334–350 (1993).
- Borst, A. & Egelhaaf, M. Dendritic processing of synaptic information by sensory interneurons. *Trends Neurosci.* **17**, 257–263 (1994).
- Borst, A. How do nerve cells compute? Dendritic integration in fly visual neurons. *Acta Physiol. Scand.* **157**, 403–407 (1996).
- Haag, J. & Borst, A. Amplification of high-frequency synaptic inputs by active dendritic membrane processes. *Nature* **379**, 639–641 (1996).
- Parks, T. N. & Rubel, E. W. Organization of the projections from n. magnocellularis to n. laminaris. *J. Comp. Neurol.* **180**, 439–448 (1975).
- Reyes, A. D., Rubel, E. W. & Spain, W. J. Membrane properties underlying the firing of neurons in the avian cochlear nucleus. *J. Neurosci.* **14**, 5352–5364 (1994).
- Reyes, A. D., Rubel, E. W. & Spain, W. J. *In vitro* analysis of optimal stimuli for phase-locking and time-delayed modulation of firing in avian nucleus laminaris neurons. *J. Neurosci.* **16**, 993–1000 (1996).
- Zhang, S. & Trussell, L. O. A characterization of excitatory postsynaptic potentials in the avian nucleus magnocellularis. *J. Neurophysiol.* **72**, 705–718 (1994).
- Manis, P. B. & Marx, S. O. Outward currents in isolated ventral cochlear nucleus neurons. *J. Neurosci.* **11**, 2865–2880 (1991).
- Agmon-Snir, H., Carr, C. E. & Rinzel, J. A simple biophysical model for analyzing phase-locking and coincidence detection in auditory neurons. *Soc. Neurosci. Abstr.* **22**, 402 (1996).
- Colburn, H. S., Han, Y. & Culotta, C. P. Coincidence model of MSO responses. *Hearing Res.* **49**, 335–346 (1990).
- Brughera, A. R., Stutman, E. R., Carney, L. H. & Colburn, H. S. A model with excitation and inhibition for cells in the medial superior olive. *Aud. Neurosci.* **2**, 219–233 (1996).
- Gerstner, W., Kempter, R., Van Hemmen, J. L. & Wagner, H. A neuronal learning rule for submillisecond temporal coding. *Nature* **383**, 76–78 (1996).
- Warchol, M. E. & Dallos, P. Neural coding in the chick cochlear nucleus. *J. Comp. Physiol.* **166**, 721–734 (1990).
- Morris, C. & Lecar, H. Voltage oscillations in the barnacle giant muscle fiber. *Biophys. J.* **35**, 193–213 (1981).
- Rinzel, J. & Ermentrout, G. B. In *Methods in Neuronal Modeling: from Synapses to Networks* (eds Koch, C. & Segev, I.) 135–169 (MIT Press, Cambridge, MA, 1989).
- Smith, P. H. Structural and functional differences distinguish principal from nonprincipal cells in the guinea pig MSO slice. *J. Neurophysiol.* **73**, 1653–1667 (1995).

Acknowledgements. This work was supported by grants from the NIH (to C.E.C.) and the Human Frontier Science Program (to H.A.-S.). We thank G. Gerstein, I. Nelken, E. W. Rubel, D. Sanes and I. Segev for comments, and the NCI Biomedical Supercomputing Center at Frederick for computer resources and technical help.

Correspondence and requests for materials should be addressed to C.E.C. (e-mail: carr@zool.umd.edu).

A prolactin-releasing peptide in the brain

Shuji Hinuma*, Yugo Habata*, Ryo Fujii*, Yuji Kawamata*, Masaki Hosoya*, Shoji Fukusumi*, Chieko Kitada*, Yoshinori Sekiguchi†, Tsutomu Kurokawa*, Osamu Nishimura†, Haruo Onda* & Masahiko Fujino*

* Discovery Research Laboratories I, Pharmaceutical Discovery Research Division, † Pharmaceutical Research Division, and ‡ Pharmaceutical Development Division, Takeda Chemical Industries Ltd, 10 Wadai, Tsukuba, Ibaraki 300-4293, Japan

Hypothalamic peptide hormones regulate the secretion of most of the anterior pituitary hormones, that is, growth hormone, follicle-stimulating hormone, luteinizing hormone, thyroid-stimulating hormone and adrenocorticotropin^{1,2}. These peptides do not regulate the secretion of prolactin^{1,2}, at least in a specific manner, however. The peptides act through specific receptors, which are referred to as seven-transmembrane-domain receptors or G-protein-coupled receptors^{3–7}. Although prolactin is important in pregnancy and lactation in mammals, and is involved in the development of the mammary glands and the promotion of milk synthesis^{8,9}, a specific prolactin-releasing hormone has remained unknown. Here we identify a potent candidate for such a hormone. We first proposed that there may still be unknown peptide hormone factors that control pituitary function through seven-transmembrane-domain receptors. We isolated the

complementary DNA encoding an 'orphan' receptor (that is, one for which the ligand is unknown). This receptor, hGR3, is specifically expressed in the human pituitary. We then searched for the hGR3 ligand in the hypothalamus and identified a new peptide, which shares no sequence similarity with known peptides and proteins, as an endogenous ligand. We show that this ligand is a potent prolactin-releasing factor for rat anterior pituitary cells; we have therefore named this peptide prolactin-releasing peptide.

We searched seven-transmembrane-domain receptors (7TMRs) with a polymerase chain reaction (PCR) method, and isolated from the human pituitary an orphan 7TMR, hGR3, that is nearly identical to GPR10 (ref. 10) and a human counterpart of rat UHR-1 (ref. 11). Quantitative analyses of UHR-1 messenger RNA

by reverse transcription PCR (RT-PCR) revealed that, of over 40 different tissues, the pituitary expressed UHR-1 mRNA at the highest level, whereas the brain, spinal cord, adrenal gland, and femur expressed UHR-1 mRNA at moderate levels. *In situ* hybridization analysis indicated that the anterior lobe abundantly expressed UHR-1 mRNA, suggesting that UHR-1 or hGR3 is particularly important in the regulation of anterior pituitary function.

On the basis of specific signal transduction in CHO cells expressing hGR3 (CHO-19P2) compared with in CHO cells transfected with a vector plasmid lacking hGR3 cDNA, we searched for an endogenous ligand of hGR3 in tissue extracts. We used several different assays for detecting signal transduction, and detected a specific response of CHO-19P2 cells to bovine hypothalamic extract when we used the arachidonic acid metabolite release assay. We used this assay as the basis for the purification of an hGR3 ligand through a combination of chromatographic procedures. The arachidonic acid metabolite-releasing activities separated into three peaks (P1, P2 and P3) with Vydac C₁₈ column chromatography. As the activity of P1 seemed to be less than those of P3 and P2, we purified P3 and P2 by Vydac diphenyl column and μ RPC C₂/C₁₈ column chromatography. As shown in Fig. 1a, two peaks of activity further separated from P3 by μ RPC C₂/C₁₈ column chromatography. Both peaks derived from P3 gave the same partial amino-terminal

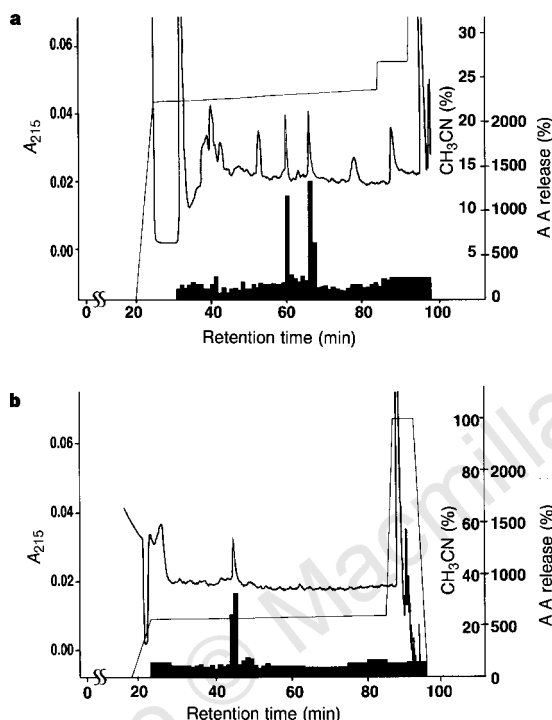


Figure 1 Purification of peptide ligands for hGR3 from bovine hypothalamic extract. The arachidonic acid metabolite (AA)-releasing activity (black areas) of each fraction is expressed as a percentage of the amount of [³H]AA released from control CHO-19P2 cells in a single assay. Thick jagged trace, A₂₁₅; thin line, percentage of CH₃CN. **a**, The profile of peptide P2 in μ RPC C₂/C₁₈ column chromatography using a SMART system. Elution was performed with a linear gradient of 22.0–23.5% CH₃CN. **b**, The profile of peptide P2 in μ RPC C₂/C₁₈ column chromatography with a linear gradient of 21.5–23.0% CH₃CN.

Bovine	1	M	K	A	V	G	A	W	L	L	C	L	L	L	L	G	L	A	L	Q	G	A	A	S	R	A	H	Q	H	S	M	30	
Rat	1	M	K	A	L	K	T	W	L	L	C	L	L	L	L	S	L	V	L	P	G	A	S	S	R	A	H	Q	H	S	M	29	
Human	1	M	K	V	L	R	A	W	L	L	C	L	L	L	M	L	G	L	A	L	R	G	A	A	S	R	T	H	R	H	S	M	30

▼

Bovine	31	E	I	R	T	P	D	I	N	P	A	W	Y	A	G	R	G	I	R	P	V	G	R	F	G	R	R	R	A	A	P	60
Rat	30	E	I	R	T	P	D	I	N	P	A	W	Y	T	G	R	G	I	R	P	V	G	R	F	G	R	R	R	A	T	P	59
Human	31	E	I	R	T	P	D	I	N	P	A	W	Y	A	S	R	G	I	R	P	V	G	R	F	G	R	R	R	A	T	L	60

↓

Bovine	61	G	D	G	P	R	P	R	R	V	P	A	C	F	R	L	E	G	G	A	E	P	S	R	A	L	P	G	R	90
Rat	60	R	D	V	T	G	L	G	L	P	R	L	S	C	L	P	L	D	G	R	A	T	K	F	S	Q	-	-	-	81
Human	61	G	D	V	P	K	P	G	L	R	P	R	L	T	C	F	P	L	E	G	A	M	S	Q	-	-	-	-	85	

▼

Bovine	91	L	T	A	Q	L	V	Q	E	98
Rat	82	-	-	-	-	-	-	R	G	83
Human	86	-	-	-	-	-	-	D	G	87

Figure 2 Amino-acid sequences of bovine, rat and human preproteins containing PrRP. These sequences were deduced from cDNAs. The filled and open arrowheads indicate the N termini of PrRP31 and PrRP20, respectively. The arrow indicates a glycine residue that is presumed to react as an amide donor. The triplets of basic amino-acid residues that constitute the typical motif of a proteolytic cleavage site are boxed with a thick line. Amino-acid residues with identical sequence in at least two of the species are boxed with a thin line.

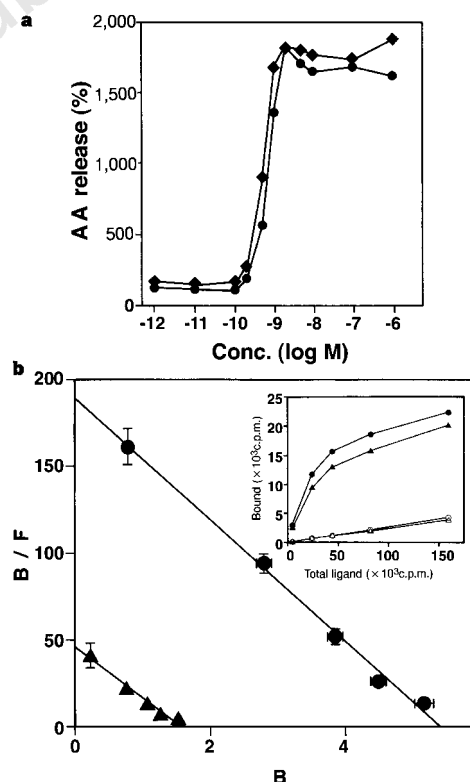


Figure 3 Specific interaction of synthetic PrRPs with hGR3 and UHR-1. **a**, Specific arachidonic acid metabolite (AA) release from CHO-19P2 cells is induced by PrRPs. CHO-19P2 cells were incubated with the indicated concentrations of bovine PrRP31 (circles) or PrRP20 (diamonds), respectively. Values represent the means of percentages of [³H]AA released in relation to a control in duplicate assays. The amount of [³H]AA released from the control was 867 c.p.m. **b**, Scatchard analysis of the binding of bovine PrRP31 to hGR3 and UHR-1. Scatchard plots of binding of [¹²⁵I]-labelled bovine PrRP31 to CHO-19P2 (circles) and CHO-UHR-1 (triangles) cells are shown. B, bound (p mol mg⁻¹ protein); F, free (nM). The inset panel represents the saturation binding data: symbols indicate the total binding of [¹²⁵I]-labelled bovine PrRP31; in the presence of 1 μ M unlabelled bovine PrRP31, to the membrane fraction of CHO-19P2 cells (filled circles) and CHO-UHR-1 cells (filled triangles), and nonspecific binding to CHO-19P2 cells (open circles) and CHO-UHR-1 cells (open triangles).

amino-acid sequence SRAHQHXMEIRTPDINPAXYAGRGIRPVG (X, unidentified residue), indicating that the difference between the two peaks might be due to a minor modification of the same peptide, such as oxidation of the methionine residue. In the case of P2, the activity was detected as a single peak (Fig. 1b). The purified P2 gave a partial N-terminal amino-acid sequence, which partly overlapped with that of P3, of TPDINPAWYAGRGIRPVGR.

We isolated bovine, rat and human cDNAs encoding the P2 and P3 peptide sequences from the brain of each species, using the purified peptide sequences as a basis for isolation of the cDNAs (Fig. 2). Although the bovine cDNA encoded a protein of 98 amino-acid residues, its N-terminal portion before Ser23 showed the typical profile of a secretory signal peptide¹², indicating that P3 was generated through cleavage of the signal sequence. The P2 peptide sequence, which starts from Thr34, indicates that the P2 peptide may be a truncated form of the P3 peptide. A typical proteolytic cleavage motif, comprising the basic amino-acid repeat Arg55, Arg56, Arg57, was conserved among the species, as was Gly54. This suggests that after cleavage between Gly54 and Arg55, Phe53 at the carboxy terminus of the predicted mature peptides might be amidated by reacting with Gly54 as an amide donor.

These results indicate that the preproprotein encoded by the bovine cDNA may generate at least two forms of mature peptide as naturally occurring endogenous ligands, that is, SRAHQHSM EIRTPDINPAWYAGRGIRPVGRF-NH₂ and TPDINPAWYAGRGIR PVGRF-NH₂, thought to correspond to the purified P3 and P2, respectively. We have named these peptides prolactin (PRL)-releas-

ing peptides (PrRPs). PrRP31 and PrRP20 correspond to peaks P3 and P2, respectively. These putative mature peptide sequences were highly conserved among the species. We analysed the tissue distribution of PrRP mRNA in rats by the quantitative RT-PCR method, and found that the medulla oblongata expressed the highest level of PrRP mRNA, whereas the hypothalamus expressed PrRP mRNA at a moderate level. However, when determined on the basis of the arachidonic acid metabolite release assay, the content of bioactive PrRP in tissue extracts was highest in the hypothalamus.

We synthesized bovine, rat and human PrRP31 and PrRP20, and examined their ability to induce arachidonic acid metabolite release from CHO-19P2 cells and CHO cells expressing UHR-1 (CHO-UHR-1 cells). Synthetic PrRP31 and PrRP20 both promoted arachidonic acid metabolite release by CHO-19P2 (Fig. 3a) and CHO-UHR-1 cells, but not by mock-transfected CHO cells. We next examined the binding of synthetic PrRPs to hGR3 and UHR-1. ¹²⁵I-labelled bovine PrRP31 specifically bound to membrane fractions prepared from both CHO-19P2 and CHO-UHR-1 cells. Competitive binding experiments showed that unlabelled PrRPs inhibited the binding of ¹²⁵I-labelled bovine PrRP31 to both CHO-19P2 and CHO-UHR-1 cells, in a dose-dependent manner. As shown in Fig. 3b, Scatchard analysis showed that both CHO-19P2 and CHO-UHR-1 cells expressed a single class of high-affinity binding site for PrRP31. The dissociation constants (*K_d*) were 2.6×10^{-11} M and 2.5×10^{-11} M, and the maximal binding sites (*B_{max}*) were 4.8 and 1.3 pmol mg⁻¹ protein, in CHO-19P2 and CHO-UHR-1 cells, respectively. However, PrRP31 with a non-

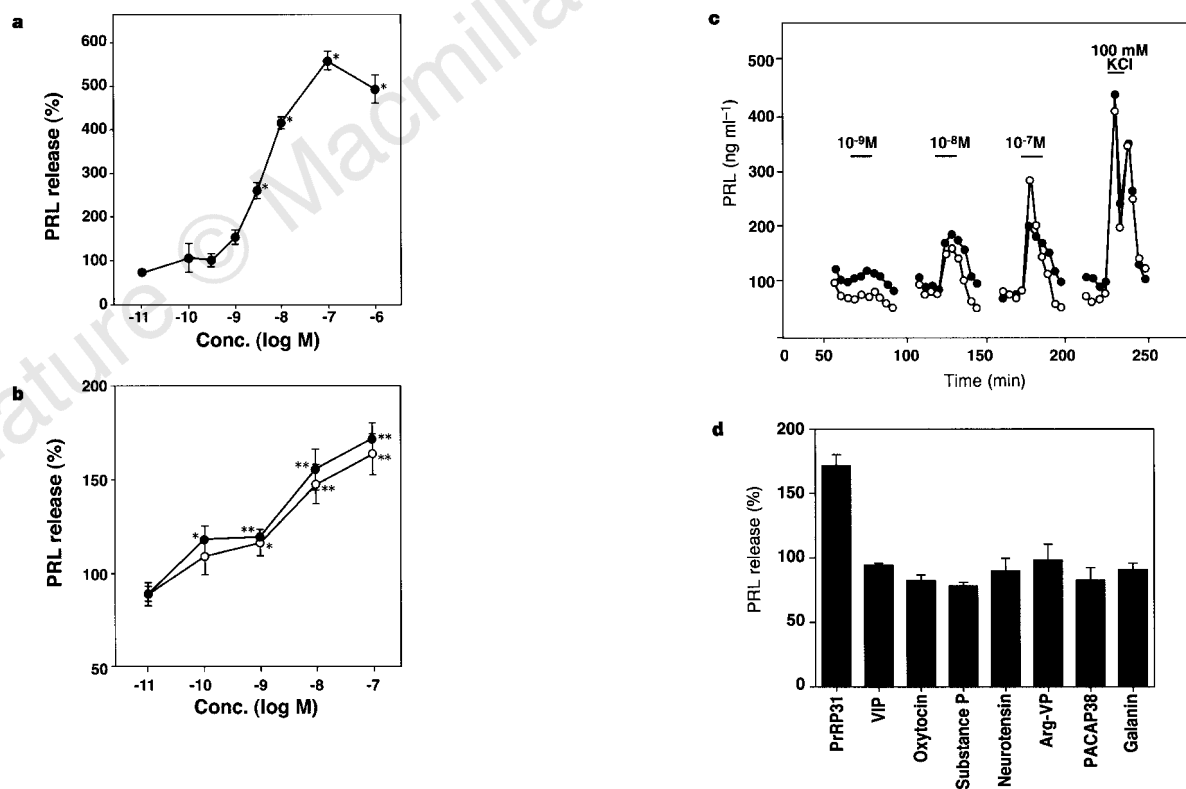


Figure 4 Promotion of PRL secretion from rat anterior pituitary cells by PrRP. **a**, Promotion of PRL secretion from RC-4B/C cells by bovine PrRP31. PRL concentrations represent the means of percentages $\pm$ s.e.m. (vertical bars) relative to a control (100%) without PrRP31, from quadruplicate assays. PRL released by the control was 1.2 ng ml^{-1} . **b**, Promotion of PRL secretion from primary cultured rat anterior pituitary cells in static incubation by bovine PrRP31. The indicated concentrations of bovine PrRP31 (filled circles) or TRH (open circles) were added to the culture. PRL concentrations are shown as for **a**. PRL released from the control was 518 ng ml^{-1} . **c**, Promotion of PRL secretion from rat

anterior pituitary cells by rat PrRP31 in a perfusion assay. Rat PrRP31 (filled circles) and TRH (open circles) at indicated doses were added to the system in different chambers at the intervals indicated by the horizontal bars. PRL concentrations taken every 3 min are indicated. **d**, Comparison of the ability of PrRP and of known peptides to promote PRL secretion in static incubation. Bovine PrRP31 and the other peptides indicated were used at 10^{-7} M. VIP, vasoactive intestinal polypeptide; Arg-VP, arginine-vasopressin; PACAP38, pituitary adenylate cyclase-activating polypeptide. Single asterisks indicate $P < 0.05$; double asterisks indicate $P < 0.01$ (when compared with control); Student's *t*-test.

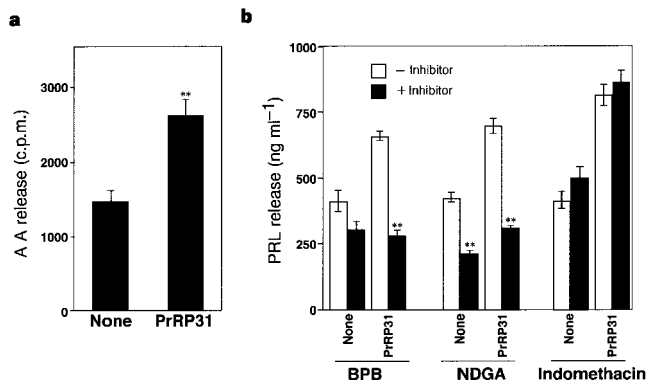


Figure 5 Relation between arachidonic acid metabolism and PRL secretion induced by PrRP in primary cultured rat anterior pituitary cells. Values represent the means of quadruplicate assays $\pm$ s.e.m. (vertical bars). Double asterisks indicate $P < 0.01$ (when compared with the control); Student's t -test. **a**, Release of arachidonic acid metabolites (AA) induced by rat PrRP31. AA release in the presence or absence (none) of rat PrRP31 at 10^{-7} M is shown. **b**, Effect of inhibitors on the secretion of PRL induced by rat PrRP31. Anterior pituitary cells untreated or treated with rat PrRP31 at 10^{-8} M were incubated in the presence (filled columns) or absence (white columns, controls) of each inhibitor at 50 μ M.

amidated carboxyl group C terminus exhibited drastically decreased arachidonic acid metabolite-releasing and receptor-binding activities towards CHO-19P2 cells, indicating that an amidated C terminus is needed for PrRPs to interact with the receptor.

To determine the physiological effects of PrRP, we first studied its effect on a rat pituitary adenoma derived cell line, RC-4B/C¹³, in which we had found the apparent expression of UHR-1 mRNA. Of the anterior pituitary hormones, only PRL was constitutively secreted at a detectable level by RC-4B/C cells under our culture conditions. Addition of PrRP31 to the culture increased the secretion of PRL from RC-4B/C cells within 30 min (Fig. 4a). We therefore expected that PrRP might act on lactotrophs to promote PRL secretion. We prepared anterior pituitary cells from lactating female rats, as the numbers of PRL-producing cells, that is, the lactotrophs, are increased in such rats⁹. We tested the effect of PrRP on hormone secretion from these cells. PrRP31 promoted PRL secretion within 15 min to 1 h of application (Fig. 4b). Thyrotropin-releasing hormone (TRH) is a potent factor that is known to be capable of promoting PRL secretion^{8,9}; PrRP31 was comparable to TRH in its potency. However, PrRP31 did not influence the secretion of the other anterior pituitary hormones, that is, growth hormone, follicle-stimulating hormone, luteinizing hormone, thyroid-stimulating hormone or adrenocorticotropin. In addition, PrRP31 immediately promoted the secretion of PRL from rat anterior pituitary cells in a perfusion assay (Fig. 4c), indicating that PrRP may stimulate lactotrophs directly to secrete PRL. Although some other known peptides, such as vasoactive intestinal polypeptide, oxytocin, substance P, neurotensin, arginine-vasopressin, pituitary adenylate cyclase-activating polypeptide and galanin, have been reported to show PRL-releasing activity *in vitro* or *in vivo*^{14–19}, these peptides did not show apparent PRL-releasing activity in primary cultured rat anterior pituitary cells, at least under our experimental conditions (Fig. 4d).

Arachidonic acid metabolism is important as a signal-transduction pathway in PRL secretion by pituitary cells²⁰, although we found that PrRP could induce Ca²⁺ influx and a partial suppression of cyclic AMP production as well as arachidonic acid metabolite release in CHO-19P2 cells. As shown in Fig. 5a, PrRP induced pronounced arachidonic acid metabolite release as well as PRL secretion even in primary cultured rat anterior pituitary cells. In addition, as shown in Fig. 5b, the specific phospholipase A₂ inhibitor BPB (4-bromophenacylbromide) and the lipoxygenase inhibitor NDGA (nordihydroguaiaretic acid) attenuated both basal and PrRP-induced PRL secretion; however, the cyclooxygenase inhibitor indomethacin had no such effect. These results show that arachidonic acid metabolism is closely linked to PrRP-induced PRL secretion, and that the lipoxygenase pathway is at least partly responsible for this effect.

The 7TMR gene family comprises many receptors which control various physiological functions. Many orphan 7TMRs have been discovered with the recent development of genome and cDNA

research. However, only a few new peptide ligands have been identified for such orphan 7TMRs. Such peptides include orphanin FQ/nociceptin^{21,22}, a snail neuro peptide²³ and the orexins²⁴. The strategy we used here would be applicable to the identification of many other unknown factors that regulate certain functions of various organs through orphan 7TMRs. The known hypothalamic hormones that regulate secretion of anterior pituitary hormones act on the pituitary through the hypophyseal portal vessel^{1,2}. As levels of biologically active PrRP were highest in the hypothalamus, PrRP might also act on the anterior pituitary through the hypothalamus/portal vessel/pituitary axis, although this is uncertain at present. Further studies will be needed to reveal the physiological significance of PrRP and its receptor *in vivo*, but we observed that the levels of expression of the mRNA for PrRP and its receptor apparently fluctuate in the medulla oblongata and pituitary during pregnancy and lactation, respectively, in rats, indicating that levels of PrRP and its receptor are closely related to the regulation of certain reproductive processes. PRL secretion is regulated in a complicated manner *in vivo* in various physiological situations^{8,9}. We assume that PrRP is important in the regulation of PRL secretion, especially in reproductive processes. PrRP and its receptor might also have functions in other tissues, including the central nervous system. Further studies of PrRP and its receptor will give us new insights into the regulatory mechanism of pituitary function and other physiological phenomena. □

Methods

Arachidonic acid metabolite release assay. We established CHO-19P2 and CHO-UHR-1 cells⁵ by transfecting CHO cells lacking the dihydrofolate reductase gene with expression vector plasmids containing hGR3 or UHR-1 cDNAs. After culturing the CHO cells in 24-well plates at 5×10^4 cells per well for 24 h, and rat anterior pituitary cells in 24-well plates at 5×10^5 cells per well for 4 days, we added [³H]arachidonic acid (NEN/Dupont) to each well at 0.25 or 2.0 μ Ci per well, and incubated the plates for a further 16 h. Then we washed the cells, added a sample, incubated the plates for 15 or 30 min, and measured the amount of [³H]arachidonic acid released into the culture supernatant.

Purification of peptide ligands for hGR3. We boiled bovine hypothalamic tissue (2 kg), homogenized it in 1 M acetic acid, and collected the supernatant. We fractionated the supernatant on a C₁₈ open column (PrepC₁₈ 125Å; Waters) with stepwise increments of 10%, 30% and 50% CH₃CN in 0.05% trifluoroacetic acid (TFA) in water. We then fractionated the 30% CH₃CN fraction on HiPrep CM-Sephacore FF (Pharmacia), with stepwise increments of 100, 200, 500 and 1,000 mM CH₃COONH₄ (pH 6.4) in 10% CH₃CN. After precipitation with acetone, the 200 mM CH₃COONH₄ fraction was serially fractionated on a Resource RPC (Pharmacia) column with a linear gradient of 15–30% CH₃CN, a Resource S (Pharmacia) column with a linear gradient of 0–0.7 M NaCl in 50 mM 2-morpholinoethanesulphonic acid (pH 5) containing 10% CH₃CN, and a Vydac C₁₈ 218TP5415 (Separations) column with a linear gradient of 20–30% CH₃CN. We further fractionated positive fractions (P3 and P2) on a Vydac diphenyl 219TP5415 (Separations) column with linear gradients of 22–25 and 21–24% CH₃CN for P3 and P2, respectively, and then

on a μ RPC C₂/C₁₈ SC 2.1/10 column (Pharmacia) in a SMART system (Pharmacia) for 60 min at a flow rate of 100 μ l min⁻¹. We analysed the N-terminal amino-acid sequences of purified peptides with a protein sequencer (model 492, ABI).

Cloning of cDNAs encoding peptide ligands for hGR3. We isolated cDNA encoding the peptide ligands from poly(A)⁺ RNA of the bovine hypothalamus by using PCR and rapid amplification of cDNA ends (RACE), with a 3'-RACE system (Gibco BRL), a Marathon cDNA amplification kit (Clontech), and degenerate primers designed on the basis of the purified peptide sequences. We isolated a rat cDNA from poly(A)⁺ RNA prepared from the dorsal region of the medulla oblongata of Wistar rats, nearly according to the strategy used for the cloning of the bovine cDNA, and then isolated a human cDNA from human brain poly(A)⁺ RNA (Clontech) with primers designed on the basis of conserved sequences found in bovine and rat cDNAs.

Synthesis of peptides. PrRPs were chemically synthesized with an automatic peptide synthesizer (model 430, ABI). They were also prepared as recombinant peptides produced in *Escherichia coli*^{25,26}. Synthetic bovine PrRP31 was labelled with ¹²⁵I by use of [¹²⁵I] Bolton-Hunter reagent (NEN/Dupont). This PrRP was used for receptor-binding assays²⁷.

Assay for hormone secretions. RC-4B/C cells were cultured at 1 × 10⁵ cells per well in 12-well tissue culture microplates for 2 days¹³. Anterior pituitary cells prepared from lactating female F344/N rats were cultured at 1.5 × 10⁵ cells per well in poly-D-lysine-coated 24-well plates (Falcon Biocoat 40414) for 4 days²⁸. The cells were then incubated with 1 ml of the medium containing each sample for 15 min to 1 h at 37 °C. Perfusion assays were performed at a flow rate of 0.33 ml min⁻¹ (ref. 29). The PRL-secreting ability of the cells was confirmed by using KCl. Amounts of pituitary hormones were determined with assay kits (follicle-stimulating hormone, thyroid-stimulating hormone, luteinizing hormone, growth hormone and PRL, Biotrak RIA/Amersham; adrenocorticotropin, Nippon DPC).

Received 26 September 1997; accepted 26 March 1998.

- Reichlin, S. Neuroendocrinology of the pituitary gland. *Toxicol. Pathol.* **17**, 250–255 (1989).
- Koenig, J. I. Pituitary gland: neurotransmitters and growth factors. *Toxicol. Pathol.* **17**, 256–265 (1989).
- Mayo, K. E., Miller, T. L., Dealmeida, V., Zheng, J. & Godfrey, P. A. The growth hormone-releasing hormone receptor: signal transduction, gene expression, and physiological function in growth regulation. *Ann. NY Acad. Sci.* **805**, 184–203 (1996).
- Kakar, S. S., Musgrove, L. C., Devor, D. C., Sellers, J. C. & Neill, J. D. Cloning, sequencing, and expression of human gonadotropin releasing hormone (GnRH) receptor. *Biochem. Biophys. Res. Commun.* **189**, 289–295 (1992).
- Hinuma, S. *et al.* Molecular cloning and functional expression of a human thyrotropin-releasing hormone (TRH) receptor gene. *Biochim. Biophys. Acta* **1219**, 251–259 (1994).
- Chalmers, D., Lovenberg, T. W., Grigoriadis, D. E., Behan, D. P. & De Souza, E. B. Corticotropin-releasing factor receptors: from molecular biology to drug design. *Trends Pharmacol. Sci.* **17**, 166–172 (1996).
- Patel, Y. C. *et al.* The somatostatin receptor family. *Life Sci.* **57**, 1249–1265 (1995).
- Frantz, A. G. Prolactin. *N. Engl. J. Med.* **298**, 201–207 (1978).
- Lamberts, S. W. & Macleod, R. M. Regulation of prolactin secretion at the level of the lactotroph. *Physiol. Rev.* **70**, 279–318 (1990).
- Marchese, A. *et al.* Cloning and chromosomal mapping of three novel genes, GPR9, GPR10, and GPR14, encoding receptors related to interleukin 8, neuropeptide Y, and somatostatin receptors. *Genomics* **29**, 335–344 (1995).
- Welch, S. K., O'Hara, B. F., Kilduff, T. S. & Heller, H. C. Sequence and tissue distribution of a candidate G-coupled receptor cloned from rat hypothalamus. *Biochem. Biophys. Res. Commun.* **209**, 606–613 (1995).
- Von Heijne, G. A new method for predicting signal sequence cleavage site. *Nucleic Acids Res.* **14**, 4683–4690 (1986).
- Hurbain-Kosmath, I. *et al.* Gonadotropes in a novel rat pituitary tumor cell line, RC-4B/C. Establishment and partial characterization of the cell line. *In Vitro Cell. Dev. Biol.* **26**, 431–440 (1990).
- Samson, W. K., Said, S. I., Snyder, G. & McCann, S. M. *In vitro* stimulation of prolactin release by vasoactive intestinal peptide. *Peptides* **1**, 325–332 (1980).
- Samson, W. K., Lumpkin, M. D. & McCann, S. M. Evidence for a physiological role for oxytocin in the control of prolactin secretion. *Endocrinology* **119**, 554–560 (1986).
- Rivier, C., Brown, M. & Vale, W. Effect of neurotensin, substance P and morphine sulfate on the secretion of prolactin and growth hormone in the rat. *Endocrinology* **100**, 751–754 (1977).
- Erfurth, E. M., Hender, P., Lundin, S. & Ekman, R. Release of prolactin as well as adrenocorticotropin after administration of arginine-vasopressin to healthy men. *Horm. Metab. Res.* **28**, 599–602 (1996).
- Hammond, P. J. *et al.* Signalling pathways mediating secretory and mitogenic responses to galanin and pituitary adenylate cyclase-activating polypeptide in the 235-1 clonal rat lactotroph cell line. *J. Neuroendocrinol.* **8**, 457–464 (1996).
- Drouhault, R. *et al.* Prolactin and growth hormone release and calcium influx are stimulated by galanin within a 'window range' of concentrations in pituitary GH3 B6. *Neuroendocrinol.* **60**, 179–184 (1994).
- Canonica, P. L., Schettini, G., Vandenegro, C. A. & MacLeod, R. M. Arachidonic acid metabolism and prolactin secretion *in vitro*: a possible role for the lipoxigenase products. *Neuroendocrinol.* **37**, 212–217 (1983).
- Reinscheid, R. K. *et al.* Orphanin FQ: a neuropeptide that activates an opioid-like G protein-coupled receptor. *Science* **270**, 792–794 (1995).
- Meunier, J. C. *et al.* Isolation and structure of the endogenous agonist of opioid receptor-like ORL1 receptor. *Nature* **377**, 532–535 (1995).

- Cox, K. J. A. *et al.* Cloning, characterization, and expression of a G-protein-coupled receptor from *Lymanaea stagnalis* and identification of a leucokinin-like peptide, PSFHSWSamide, as its endogenous ligand. *J. Neurosci.* **17**, 1197–1205 (1997).
- Sakurai, T. *et al.* Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell* **92**, 575–585 (1998).
- Koyama, N., Kuriyama, M., Amano, N. & Nishimura, O. A novel procedure for the preparation of biologically active recombinant peptides using a cyanation reaction. *J. Biotechnol.* **32**, 273–281 (1994).
- Nakagawa, S. *et al.* Chemical cleavage of recombinant fusion protein to yield peptide amides. *J. Am. Chem. Soc.* **116**, 5513–5514 (1994).
- Miyamoto, Y. *et al.* Cloning and expression of a complementary DNA encoding the bovine receptor for pituitary adenylate cyclase-activating polypeptide (PACAP). *Biochim. Biophys. Acta* **1218**, 297–307 (1994).
- Shiota, K. *et al.* *In vitro* thyrotropin release with thyrotropin-releasing hormone and an analogue, DN-1417. *Acta Endocrinol.* **106**, 71–78 (1984).
- Hyde, J. E., Murai, I. & Ben-Jonathan, N. The rat posterior pituitary contains a potent prolactin-releasing factor: studies with perfused anterior pituitary cells. *Endocrinology* **121**, 1531–1539 (1987).

Acknowledgements. We thank H. Okazaki and K. Tsukamoto for discussions throughout this study, and Y. Ishibashi, J. Noguchi, Y. Matsumoto, T. Moriya and M. Suenaga for collaboration.

Correspondence and requests for materials should be addressed to S.H. (e-mail: Hinuma_Shuji@takeda.co.jp).

Elastin is an essential determinant of arterial morphogenesis

Dean Y. Li^{††}, Benjamin Brooke[†], Elaine C. Davis[‡], Robert P. Mecham[§], Lise K. Sorensen^{||}, Beth B. Boak^{||}, Ernst Eichwald[†] & Mark T. Keating^{*†||}

^{*} Cardiology Division and [†] Department of Pathology, University of Utah Health Sciences Center, Salt Lake City, Utah 84112, USA

[†] Department of Human Genetics, Eccles Institute of Human Genetics and

^{||} Howard Hughes Medical Institute, Salt Lake City, Utah 84112-5330, USA

[‡] Department of Cell Biology and Neuroscience, University of Texas Southwestern Medical Center, Dallas, Texas 75235-9039, USA

[§] Department of Cell Biology and Physiology and Department of Medicine, Washington University School of Medicine, St Louis, Missouri 63110, USA

Elastin, the main component of the extracellular matrix of arteries, was thought to have a purely structural role¹. Disruption of elastin was believed to lead to dissection of arteries^{2,3}, but we showed that mutations in one allele encoding elastin cause a human disease in which arteries are blocked, namely, supravalvular aortic stenosis^{4,5}. Here we define the role of elastin in arterial development and disease by generating mice that lack elastin. These mice die of an obstructive arterial disease, which results from subendothelial cell proliferation and reorganization of smooth muscle. These cellular changes are similar to those seen in atherosclerosis. However, lack of elastin is not associated with endothelial damage, thrombosis or inflammation, which occur in models of atherosclerosis. Haemodynamic stress is not associated with arterial obstruction in these mice either, as the disease still occurred in arteries that were isolated in organ culture and therefore not subject to haemodynamic stress. Disruption of elastin is enough to induce subendothelial proliferation of smooth muscle and may contribute to obstructive arterial disease. Thus, elastin has an unanticipated regulatory function during arterial development, controlling proliferation of smooth muscle and stabilizing arterial structure.

We isolated murine genomic clones encoding elastin (ELN) from a SV129 λ FIX II library using an ELN complementary DNA clone. A targeting vector was constructed to delete 4.0 kilobases (kb) of the promoter and exon 1, resulting in a null mutation (Fig. 1a). The vector was electroporated into R1 embryonic stem cells and homologous recombinants were isolated by positive-negative selection⁶. Of 160 clones, we identified 3 as homologous recombinants by Southern blot analysis (Fig. 1b). These three clones were micro-injected into C57BL/6 blastocysts and implanted in pseudopregnant

females⁷. All three independent cell lines generated chimaeric mice that transmitted the null *ELN* allele. The resulting *ELN*^{+/-} mice were mated to generate *ELN*^{-/-} mice.

The disrupted *ELN* allele is a null allele, as shown by the absence of *ELN* messenger RNA and protein in *ELN*^{-/-} mice (Fig. 1c-f). Elastin expression during fetal development is largely confined to the vascular system and begins in the last third of gestation¹. *In situ* hybridization showed the absence of *ELN* mRNA in the aortae of *ELN*^{-/-} mice and its presence in *ELN*^{+/+} mice (Fig. 1c, d). Northern blot analysis showed an absence of *ELN* mRNA in *ELN*^{-/-} mice at birth (0.5 days post partum (P); data not shown). Hart staining showed the absence of elastin protein in *ELN*^{-/-} mice (Fig. 1e, f).

ELN^{-/-} mice survived gestation but died by P4.5. To determine the consequences of elastin deficiency, we studied several developmental stages of the mice (Fig. 2). The histological appearances of ascending aortae from *ELN*^{-/-} and *ELN*^{+/+} mice were indistinguish-

able until embryonic day (E) 17.5. But from E17.5, the outer and inner diameters of the aorta in *ELN*^{-/-} mice became progressively smaller. In *ELN*^{+/+} mice, in contrast, the aorta continued to expand. The diameter of the arterial wall in *ELN*^{-/-} mice became progressively thicker after E17.5. This change was caused by subendothelial accumulation of cells, a process that eventually obliterated the vascular lumen. Cell counts of the ascending aorta showed a marked difference between the number of cells in *ELN*^{+/+} and *ELN*^{-/-} mice at E17.5 ($P < 0.025$; $n = 5$). By P2.5, there were 76% more cells in aortic cross-sections of *ELN*^{-/-} mice than in *ELN*^{+/+} mice ($P < 0.005$; $n = 5$).

The aorta and other large, conduit arteries are often referred to as elastic arteries because of the high elastin content and the highly organized structure of elastic lamella⁸. In distributing or muscular arteries, however, the relative proportion of elastin declines and the structure of elastic fibres in the medium is less defined. To determine the consequences of elastin deficiency in different arteries, we

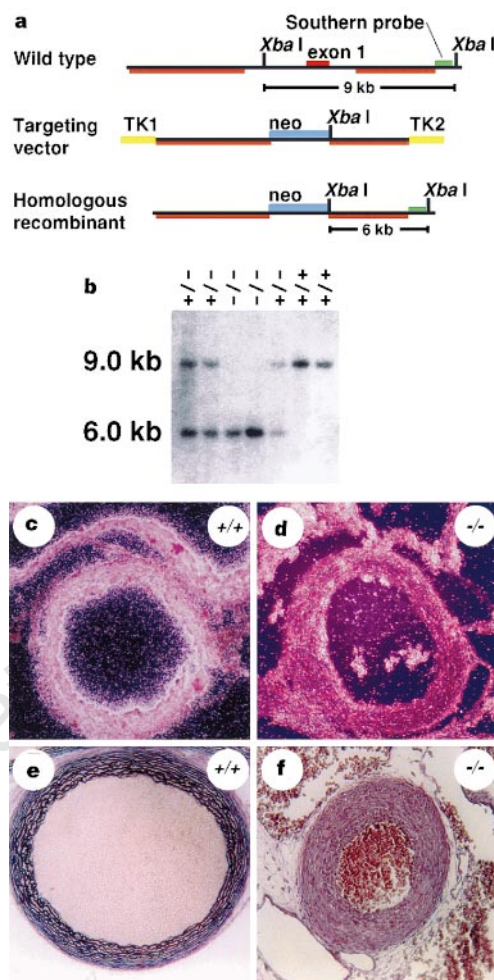


Figure 1 Disruption of the murine *ELN* gene. **a**, Diagram of the targeting vector designed to delete the promoter and exon 1 of the *ELN* gene. TK, thymidine kinase. **b**, Southern blot analysis of tail DNA, digested with *Xba*I, from *ELN*^{+/+}, *ELN*^{+/-}, and *ELN*^{-/-} mice. The 9-kb wild-type and 6-kb targeted *Xba*I fragments were identified using a probe outside the 3' region of homology. **c, d**, *In situ* hybridization of *ELN*^{+/+} (**c**) and *ELN*^{-/-} (**d**) mice at E17.5 was used to study *ELN* expression. Dark-field photomicrographs show the presence of silver grains in *ELN*^{+/+} mice and their absence in *ELN*^{-/-} mice. The results demonstrate that no *ELN* mRNA is detected in *ELN*^{-/-} mice. **e, f**, Hart stain detected elastic fibres (stained dark brown) in ascending aortae from *ELN*^{+/+} mice (**e**) at birth (P0.5). Elastin is organized into polymers that form concentric rings of elastic lamella around the arterial lumen. Each elastic lamella alternates with a ring of smooth muscle, forming a lamellar unit. There were no elastic fibres in *ELN*^{-/-} mice (**f**).

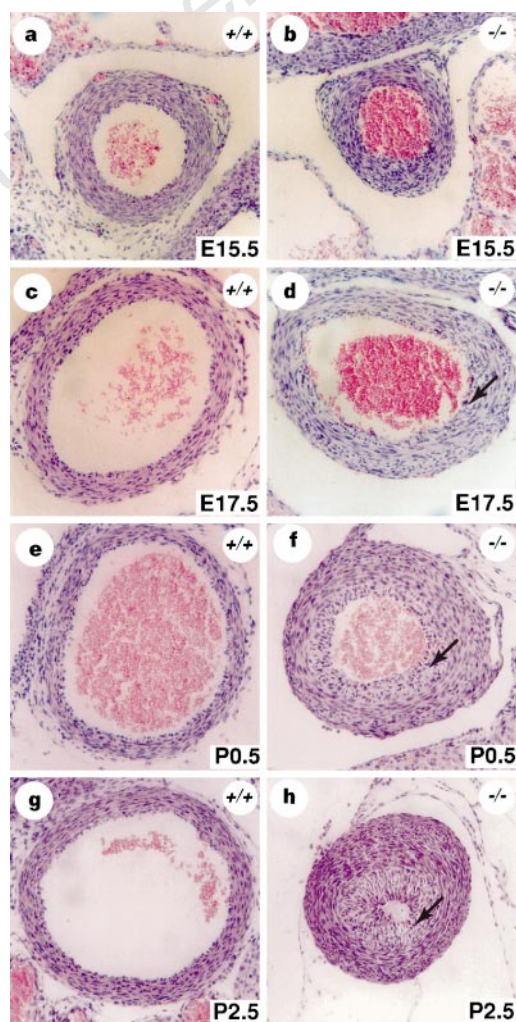


Figure 2 Histology of different developmental stages in *ELN*^{+/+} and *ELN*^{-/-} aortae. **a-h**, Cross-sections of ascending aortae (at the level of the pulmonary artery) stained with haematoxylin and eosin were examined at E15.5 (**a, b**), E17.5 (**c, d**), P0.5 (**e, f**) and P2.5 (**g, h**). No difference was noted at E15.5. Subsequent timepoints revealed subendothelial accumulation of cells in *ELN*^{-/-} mice (arrows) and obliteration of the lumen by P2.5. All *ELN*^{-/-} mice were dead by P4.5, presumably from the rapid obliteration of the aorta and other arteries. The ratios of lung weight to body weight and of heart weight to body weight in *ELN*^{-/-} and *ELN*^{+/+} mice were comparable, suggesting that pulmonary oedema was not the cause of death ($n = 8$; data not shown). There was no evidence of cardiac hypertrophy, dilatation or ischaemia.

examined a pulmonary arteriole, the internal mammary artery, a distal portion of the subclavian artery, and the abdominal aorta in $ELN^{+/+}$ and $ELN^{-/-}$ mice (Fig. 3). We found increased cellularity and reduced luminal diameter in systemic and pulmonary arteries of all sizes, including distributing arteries and arterioles (Fig. 3a–d). These data indicate that development of several different arteries is abnormal in mice lacking elastin.

The accumulating cells in aortic cross-sections of $ELN^{-/-}$ mice were morphologically distinct from cells in the smooth muscle of normal arteries. To identify these cells, we immunostained aortic sections with antisera against smooth-muscle α -actin. The accumulating subendothelial cells present in $ELN^{-/-}$ aortae stained positively (Fig. 4a, b). In sagittal sections, these cells were longitudinally orientated (Fig. 4c, d). In contrast, smooth muscle cells in the media of control animals were circumferentially orientated. Thus, obliteration of arterial lumens in $ELN^{-/-}$ aortae results from subendothelial accumulation and reorientation of smooth muscle.

To examine the mechanism underlying accumulation of smooth muscle in $ELN^{-/-}$ mice, we stained aortic sections with antisera against proliferating-cell nuclear antigen (PCNA; Fig. 4e, f). At E17.5, the percentage of cells staining positive for PCNA was greater in $ELN^{-/-}$ mice than in $ELN^{+/+}$ mice (88% versus 35%). Staining of $ELN^{-/-}$ aortae was increased throughout the aortae, but there was a gradient of staining, with subendothelial cells showing prominent

staining. We used intestinal sections as controls; PCNA was present in the crypts and absent in the villi of these sections, as expected (data not shown). These results indicate that subendothelial proliferation of smooth muscle is a mechanism underlying the arterial pathology observed in $ELN^{-/-}$ mice.

Obstructive arterial diseases, such as atherosclerosis and restenosis, are the main cause of morbidity and mortality in industrialized nations⁹. Although much is known about vascular risk factors such as hypercholesterolaemia, hypertension, diabetes and cigarette abuse, the pathogenic mechanism through which these diverse factors converge to give a common pathology is unknown^{10,11}. The hallmark of obstructive arterial disease is subendothelial accumulation of smooth-muscle cells^{11,12}. Recurrent endothelial injury, thrombosis and inflammation are thought to

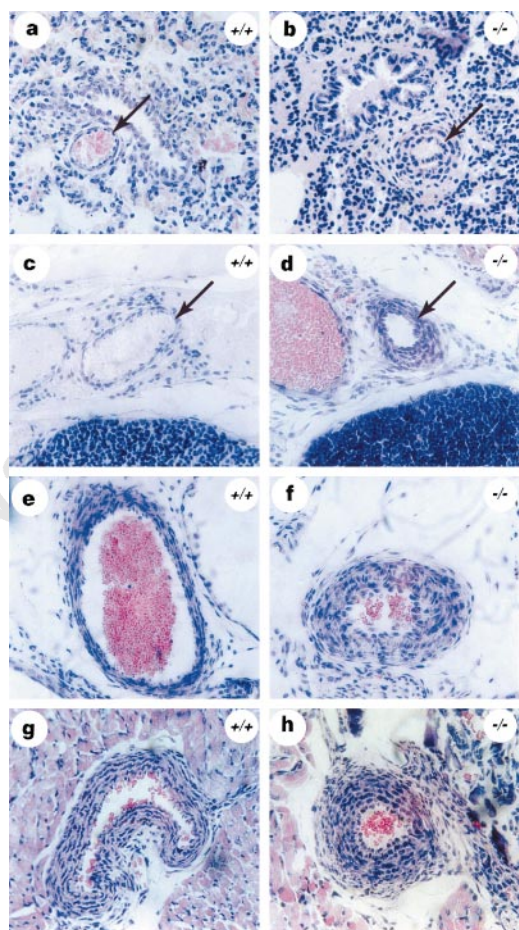


Figure 3 Consequences of elastin deficiency in different arteries. **a–h**, Sections of pulmonary arterioles (arrow) at the level of a terminal bronchiole at P2.5 (**a, b**), the internal mammary artery (arrow) at P2.5 (**c, d**), a distal portion of the subclavian artery at P0.5 (**e, f**), and the abdominal aorta at P0.5 (**g, h**) were stained with haematoxylin and eosin and examined. There is increased cellularity and narrowing of the lumen all of these $ELN^{-/-}$ arteries.

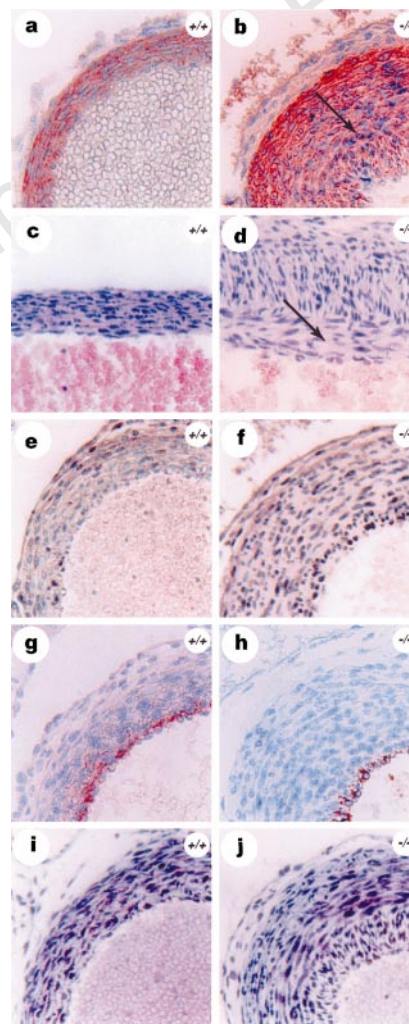


Figure 4 Mechanism of luminal obliteration in $ELN^{-/-}$ mice. **a, b**, Sections of mice at P0.5 were immunostained with antisera against smooth-muscle α -actin. The subendothelial cells in $ELN^{-/-}$ mice stained positively (arrow), indicating that they were smooth muscle. **c, d**, Sagittal sections of the ascending aorta at P2.5, stained with haematoxylin and eosin. The subendothelial cells in $ELN^{-/-}$ mice were longitudinally orientated (arrow), whereas normal orientation is circumferential. **e, f**, Aortic sections of E17.5 mice were immunostained with antisera against PCNA (dark brown nuclei). The percentage of cells stained with PCNA was greater in $ELN^{-/-}$ mice than in $ELN^{+/+}$ mice (88% versus 35%), indicating increased cell proliferation in $ELN^{-/-}$ aortae. **g, h**, Sections of mice at P0.5 were immunostained with antisera against von Willebrand factor (red). There was no evidence of endothelial damage. **i, j**, Sections of mice at P0.5 were incubated with Masson trichrome, which stains collagen green. There was no collagen deposition.

induce this pathology¹². To determine whether endothelial damage was responsible for accumulation of smooth muscle in $ELN^{-/-}$ aortae, we used histochemistry, immunohistochemistry, and electron microscopy to examine aortic sections. Sagittal and cross-sections of aortae stained with antisera against smooth-muscle α -actin revealed a single uninterrupted layer of unstained cells, the endothelium. The identity of these cells was further defined using antisera against von Willebrand factor (Fig. 4g, h). There was no evidence of endothelial damage or disruption. Immunohistochemical analysis showed no evidence of endothelial activation and there was no ultrastructural evidence of rounding, protrusions, detachment or apoptosis of endothelial cells (data not shown). These results, however, cannot exclude the possibility of endothelial dysfunction.

To determine whether inflammation was responsible for cellular pathology in $ELN^{-/-}$ aortae, we examined sections with hematoxylin and eosin (Figs 2, 3). There was no inflammation. Immunohistochemical analyses using the leukocyte-specific antibody Mac-3 also failed to show inflammation (data not shown). Gross and histological examination of arteries showed no evidence of thrombosis. As collagen is a ligand for integrin receptors and can modulate proliferation of smooth muscle, we looked for increased collagen levels using Masson trichrome¹³; no increase was found (Fig. 4i, j). Use of electron microscopy confirmed the lack of inflammation and

the normal distribution of other extracellular matrix proteins, including collagen.

Hypertension and haemodynamic stress are important mediators of obstructive arterial disease¹⁴. To determine whether stress to the aortic walls was responsible for increased cell proliferation in $ELN^{-/-}$ aortae, we took aortic segments from $ELN^{+/+}$ and $ELN^{-/-}$ embryos at E16.5 and incubated them in organ culture (Fig. 5). In this system, there was no blood flow and no haemodynamic stress. Luminal obliteration by proliferating smooth-muscle cells occurred in $ELN^{-/-}$ mice within 24 h, but not in $ELN^{+/+}$ mice. The proliferative index for $ELN^{-/-}$ aortae in organ culture was markedly increased compared with controls (27% versus 14%; $n = 8$). These indices were similar to those seen for $ELN^{-/-}$ and $ELN^{+/+}$ aortae between E15.5 and E17.5 *in vivo* (25% versus 10%; $n = 5$). Thus, blood flow, haemodynamic stress, gross endothelial damage, thrombosis, inflammation, and fibrosis are not responsible for subendothelial accumulation of smooth muscle in $ELN^{-/-}$ mice and therefore are not necessary for inducing these pathological changes. Instead, ELN disruption is enough to initiate accumulation and reorganization of smooth muscle. We propose that endothelial injury, thrombosis, and inflammation may contribute to obstructive arterial pathology by disrupting elastin.

Our results indicate a regulatory function for elastin and define a fourth stage of arterial development. In this stage, the extracellular matrix matures, smooth-muscle cells exit the cell cycle and become organized, and the artery stabilizes^{15–18}. Arterial development in mice lacking elastin was indistinguishable from controls until E17.5, indicating that elastin has little or no effect on the first three stages of vascular development¹⁶. During subsequent development, however, $ELN^{-/-}$ aortae became smaller and thicker, arterial diameter declined and the arterial lumen was eventually obliterated. The cellular mechanism underlying these changes was subendothelial accumulation of arterial smooth muscle, a process that involved cell proliferation, migration and reorganization. Thus, elastin is a molecular determinant of late arterial morphogenesis, stabilizing arterial structure by regulating proliferation and organization of vascular smooth muscle. □

Methods

Construction of the targeting vector. Two phages from a λ FIX II library (Stratagene), encompassing exon 1 and spanning a total of 26 kb, were isolated, mapped and subcloned. To construct a targeting vector, we used a 6.5-kb *Xba*I fragment as the 5' region of homology and a 4.2-kb *Bam*HI–*Hind*III fragment as the 3' region of homology. Positive selection was provided by a 3.1-kb *Sac*I–*Cl*aI fragment containing the neomycin-resistance gene under the control of the RNA polymerase II promoter. These fragments were cloned into the vector TK1–TK2A, which provided two thymidine kinase genes for negative selection¹⁹.

Generation of mice. Culture, selection of embryonic stem cells, and screening of targeted clones were done as described^{6,7}. The R1 embryonic stem cell line used was derived from strain 129/SV $\times$ 129/SVJ. Clones were screened by Southern blot analysis of *Xba*I-digested genomic DNA, probed with a 3.0-kb *Hind*III–*Xba*I fragment. Random integration events were excluded using a probe derived from the 3' region of homology. The homologous recombinant clones were injected into blastocysts of C57B1/6J mice and transferred into uteri of pseudopregnant C57B1/6J females. The resulting chimaeric animals were backcrossed to C57B1/6J mice, and heterozygous mutants were identified by genomic Southern blotting of tail DNA. Heterozygous mice were mated for three generations with C57B1/6J mice and brother–sister matings were carried out to generate homozygous mutants.

Northern blotting and *in situ* hybridization. Poly(A)⁺ RNA was extracted from the visceral organs of the thorax from mice at P0.5 using a Micro-FastTrack Kit (Invitrogen). RNA was separated by electrophoresis on a 1.0% denaturing agarose gel, transferred to a Hybond filter (Amersham), and hybridized with a ³²P-labelled 0.85-kb fragment of mouse *ELN* cDNA. Filters were rehybridized with a 1.5-kb fragment of human cardiac actin cDNA.

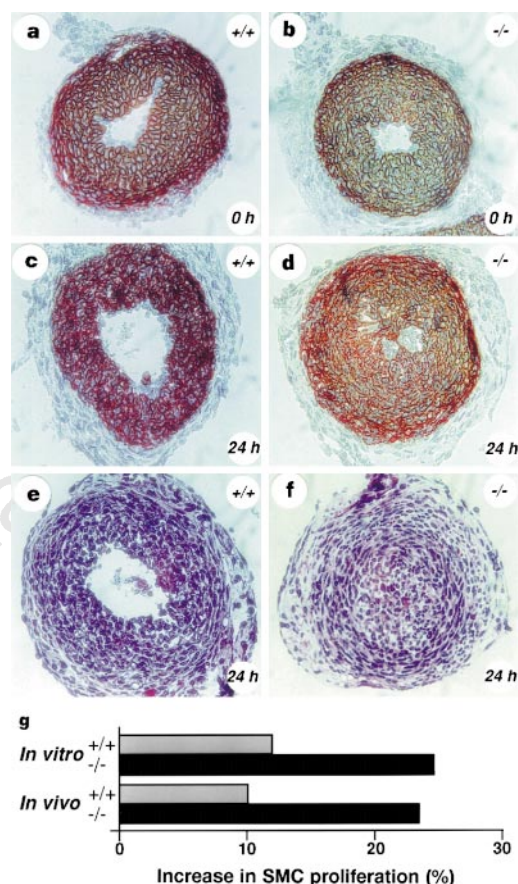


Figure 5 Organ culture of $ELN^{+/+}$ and $ELN^{-/-}$ ascending aortae. **a–f**, Cross-sections of $ELN^{+/+}$ (**a**, **c**, **e**) and $ELN^{-/-}$ (**b**, **d**, **f**) aortae with no incubation (**a**, **b**) and 24 h of incubation (**c–f**) in DMEM plus 5% fetal bovine serum. Sections in **a–d** were stained with antisera against smooth-muscle α -actin whereas sections in **e**, **f** were stained with haematoxylin and eosin. **d**, **f**, Obliteration of the $ELN^{-/-}$ aortae occurs after 24 h of incubation. **g**, Comparison of *in vitro* and *in vivo* proliferative indices. The proliferative indices of $ELN^{+/+}$ and $ELN^{-/-}$ aortae in organ culture (E16.5, 24-h incubation; $n = 8$) were comparable to those obtained *in vivo* (E15.5–E17.5; $n = 5$).

Paraffin-embedded cross-sections were hybridized with specific ^{35}S -labelled antisense and sense *ELN* complementary RNA probes. Probes were synthesized from linearized templates using T3 and T7 RNA polymerases. Hybridization and washes were performed as described²⁰. Sections were exposed to emulsion for several weeks and analysed using bright- and dark-field microscopy.

Histological examination. Mice were fixed overnight in either 4% paraformaldehyde or methyl Carnoy's at 4°C and embedded in paraffin. Transverse sections of the ascending and descending aortae were examined at the point at which the pulmonary artery crosses the ascending aorta posteriorly and bifurcates. Sections were stained with haematoxylin and eosin, Hart, and Masson trichrome. Cell numbers were counted on two separate occasions by people blinded to the genotype. The statistical significance was calculated by comparison of the means using Student's *t*-test analysis.

Immunohistochemistry. Tissue samples were immunostained with a monoclonal antibody against smooth-muscle α -actin (clone 1A4, 1:400; Sigma), an antibody against human von Willebrand factor (1:1,000; DAKO), or an anti-PCNA monoclonal antibody (clone PC10, 3 $\mu\text{g ml}^{-1}$; Calbiochem).

Biotinylated donkey anti-rabbit antibody (RPN 1004, 1:2,000; Amersham) was used as a secondary antibody for von Willebrand factor, and biotinylated goat anti-mouse IgG2a antibody (RPN 1181, 1:100; Amersham) was used as a secondary antibody for smooth-muscle α -actin and PCNA. Sections immunostained for smooth-muscle α -actin or von Willebrand factor were developed with 3-amino-9-ethylcarbazole chromagen (DAKO) and counterstained with Mayer's haematoxylin (Sigma). Sections immunostained for PCNA were developed in 3,3'-diaminobenzidine chromagen (Vector Laboratories) and counterstained with methyl green.

Organ culture. Ascending aortae were dissected from mouse embryos at E16.5 and washed in PBS. Aortae were cultured for 24 h at 37°C in 5% fetal bovine serum in DMEM. After 24 h of culturing, each aorta was fixed in 1 ml methyl Carnoy's fixative for 16 h. *ELN*^{+/+} and *ELN*^{-/-} aortae were selected for processing and embedding in paraffin.

Received 20 October 1997; accepted 9 March 1998.

1. Parks, W. C., Pierce, R. A., Lee, K. A. & Mecham, R. P. in *Advances in Molecular and Cell Biology* (Kleinman, H. K. ed.) Vol. 6, 133–182 (JAI, Greenwich, 1993).
2. Thompson, R. W. Basic science of abdominal aortic aneurysms: emerging therapeutic strategies for an unresolved clinical problem. *Curr. Opin. Cardiol.* **11**, 504–518 (1996).
3. Terpin, T. & Roach, M. R. A biophysical and histological analysis of factors that lead to aortic rupture in normal and lathyrotic turkeys. *Can. J. Phys. Pharmacol.* **65**, 395–400 (1987).
4. Curran, M. et al. The elastin gene is disrupted by a translocation associated with supravalvular aortic stenosis. *Cell* **73**, 159–168 (1993).
5. Ewart, A. et al. Hemizygosity at the elastin locus in a developmental disorder. *Nature Genet.* **5**, 11–16 (1993).
6. Mansour, S. L., Thomas, K. R. & Capecchi, M. R. Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348–352 (1988).
7. Thomas, K. R. & Capecchi, M. R. Targeted disruption of the murine *int-1* proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. *Nature* **346**, 847–850 (1990).
8. Bloom, W. & Fawcett, D. *A Textbook of Histology* 10th edn, 398–402 (Saunders, Philadelphia, 1975).
9. American Heart Association *Heart and Stroke Facts: 1996 Statistical Supplement* (American Heart Association, Dallas, 1996).
10. Schwartz, S. M., deBlois, D. & O'Brien, E. R. M. The intima: soil for atherosclerosis and restenosis. *Circulation* **77**, 445–465 (1995).
11. Gibbons, G. H. & Dzau, V. J. Molecular therapies for vascular diseases. *Science* **272**, 689–693 (1996).
12. Ross, R. The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature* **362**, 801–809 (1993).
13. Koyama, H., Raines, E. W., Bornfeldt, K. E., Roberts, J. M. & Ross, R. Fibrillar collagen inhibits arterial smooth muscle proliferation through regulation of cdk2 inhibitor. *Cell* **87**, 1069–1078 (1996).
14. Langille, B. C. & Ojhu, M. Blood flow dynamics, atherosclerosis and bypass graft failure. *Trends Cardiovasc. Med.* **7**, 111–118 (1997).
15. Hanahan, D. Signaling vascular morphogenesis and maintenance. *Science* **277**, 48–50 (1997).
16. Folkman, J. & D'Amore, P. D. Blood vessel formation: what is its molecular basis? *Cell* **87**, 1153–1155 (1996).
17. Schwartz, S. M., Heimark, R. L. & Majesky, M. W. Developmental mechanism underlying pathology of arteries. *Physiol. Rev.* **70**, 1177–1209 (1990).
18. Owens, G. K. Regulation of differentiation of vascular smooth muscle cells. *Physiol. Rev.* **75**, 487–517 (1995).
19. Deng, C., Thomas, K. R. & Capecchi, M. R. Location of crossovers during gene targeting with insertion and replacement vectors. *Mol. Cell. Biol.* **13**, 2134–2140 (1993).
20. Prosser, I. W. et al. Regional heterogeneity of elastin and collagen gene expression in intralobar arteries in response to hypoxic pulmonary hypertension as demonstrated by *in situ* hybridization. *Am. J. Pathol.* **135**, 1073–1088 (1989).

Acknowledgements. We thank D. Atkinson for help with manuscript preparation; M. Parmacek, K. Thomas, R. White, S. Prescott, J. Kaplan, G. Zimmerman and M. Sanguinetti for comments; and K. Thomas for technical advice. ES cells and chimaeric mice were generated in M. R. Capecchi's laboratory. This work was supported by the NIH and by the Bristol-Myers Squibb Foundation.

Correspondence and requests for materials should be addressed to D.Y.L. (dean@howard.genetics.utah.edu).

Protein-primed RNA synthesis by purified poliovirus RNA polymerase

Aniko V. Paul, Jacques H. van Boom*, Dmitri Filippov* & Eckard Wimmer

Department of Molecular Genetics and Microbiology, School of Medicine, Health Science Center, State University of New York, Stony Brook, New York 11794–5222, USA

* Leiden University, Gorlaeus Laboratory, PO Box 9502, 2300 RA Leiden, The Netherlands

A small protein, VPg, is covalently linked to the 5' end of the plus-stranded poliovirus genomic RNA^{1–3}. Poliovirus messenger RNA, identical in nucleotide sequence to genomic RNA, is not capped at its 5' end by the methylated structure that is common to most eukaryotic mRNAs. These discoveries presented two problems. First, as cap structures are usually required for translation of mRNA into protein, how does this uncapped viral RNA act as a template for translation? Second, what is the function of VPg? The identification of the internal ribosomal-entry site, which allows the entry of ribosomes into viral mRNA independently of the 5' mRNA end, has solved the first conundrum^{4–6}. Here we describe the resolution of the second problem. VPg is linked to the genomic RNA through the 5'-terminal uridylic acid of the RNA. We show that VPg can be uridylylated by the poliovirus RNA

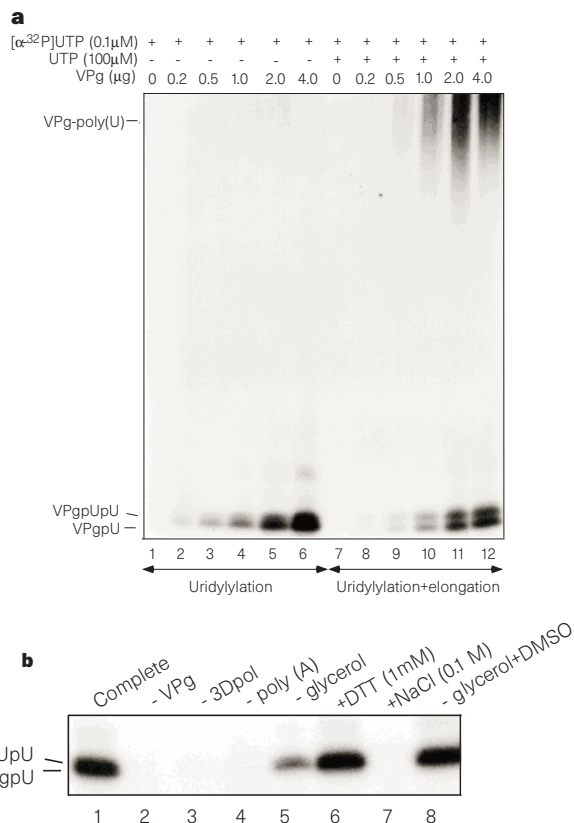


Figure 1 Uridylation of VPg and synthesis of VPg-poly(U) *in vitro* by poliovirus polymerase 3D^{pol}. **a**, Synthesis of VPgUpU and VPg-poly(U) as a function of VPg concentration. **b**, Requirements for uridylation of VPg. All samples contained [α-³²P]UTP (0.1 μM); different reaction components were omitted or added as indicated. DMSO, dimethyl sulphoxide.

polymerase 3D^{pol}. Uridylylated VPg can then prime the transcription of polyadenylate RNA by 3D^{pol} to produce VPg-linked poly(U). Initiation of transcription of the poliovirus genome from the polyadenylated 3' end therefore depends on VPg.

VPg is a protein of only 22 amino acids⁷ whose single tyrosine residue provides the link to the 5'-terminal uridylylic acid of the poliovirus genome^{8,9}. Biochemical and genetic data have indicated that protein(s) mapping to the so-called 3D region of the viral polyprotein are involved in forming the O⁴(5'-uridylyl)tyrosine bond¹⁰⁻¹². Therefore, we incubated synthetic VPg with purified 3D^{pol}, the primer-dependent RNA transcriptase component of the polyprotein¹³, and [α -³²P]UTP. No products were formed unless polyadenylate (poly(A)) was added to the reaction mixture. 3D^{pol} then synthesized two products with characteristics of VPgpU and VPgpUpU¹⁰⁻¹², which are referred to here as VPgpU(pU) (Fig. 1a, lanes 1-6; Fig. 1b). Digestion of ³²P-labelled VPgpU(pU) with HCl and analysis of the products by thin-layer chromatography or paper electrophoresis showed that the tyrosine residue of VPg was uridylylated (data not shown). Glycerol (6%) or dimethyl sulphoxide (5%) were, surprisingly, strong stimulants for uridylylation (Fig. 1b, lanes 5, 8), whereas 100 mM NaCl inhibited the reaction (lane 7) and dithiothreitol (DTT) had no effect (lane 6).

Nucleotidylation of VPg was specific as it required specifically UTP and poly(A). None of the other three ribonucleoside triphosphates was able to replace UTP, regardless of the template used ((poly(G), poly(C) or poly(U)), and no label was transferred to VPg from [γ -³²P]ATP (data not shown). These findings coincide with the observation that the 3'-terminal poly(A) of the poliovirus genome¹⁴ is the site at which genome replication commences¹⁵⁻¹⁸.

The yield of VPgpU(pU) was proportional to the concentration of the precursor, VPg. Products of higher relative molecular mass indicative of transcription of the poly(A) template were not apparent (Fig. 1a, lanes 1-6). Polymers were synthesized, however, when the UTP concentration was increased to >10 μ M (Fig. 1a, lanes 10-12; Fig. 2, lanes 3-5). Synthesis of these products was strictly dependent upon the presence of VPg (Fig. 2, lanes 6, 7), indicating that transcription of the poly(A) template, to produce poly(U), may

be primed by VPg. Analysis of the homopolymeric RNA product formed in the elongation reaction revealed chains of varying length (Fig. 3, lane 1). The product was partially hybridized to the poly(A) template: it was partially sensitive to either of the single-strand-specific RNases A or T₂, or to double-strand-specific RNase V₁, but it was completely digested with mixtures of RNases A and V₁ (Fig. 3, lane 3) or T₂ and V₁ (data not shown). After removal of VPgpU(pU) from the reaction mixture (Fig. 3, lane 1), treatment of the homopolymer VPg-poly(U) with RNases A and V₁ yielded VPgpUpU (lane 3). This, in turn, was cleaved to VPgpU with alkaline phosphatase (lane 4). Phosphorimager analysis revealed the expected values of 1% and 0.5% of the total label in VPgpUpU and VPgpU, respectively. This finding confirmed that all homopolymer chains were linked to VPg.

Neither uridylylation of VPg nor elongation of VPgpU(pU) into VPg-poly(U) was affected by 0.1% non-ionic detergent N-P40 (data not shown). However, the yield of products in a standard reaction was low (see Methods), as was the case in a similar reaction involving nucleotidylation of the phage ϕ 29 terminal protein by the phage-specific DNA polymerase (see ref. 19 and refs therein). The reason for the low yield of VPgpU(pU) or VPg-poly(U) is not known. It is possible, and even likely, that efficient uridylylation may require extra specific viral or cellular factors, and/or a hydrophobic (membranous?) environment to stabilize protein-RNA complexes.

Poliovirus 3D^{pol} oligomerizes with increasing molar concentration, thereby improving its ability to bind to RNA and use RNA primers to initiate RNA synthesis²⁰. We have found that uridylylation and elongation are optimal at 1 μ M 3D^{pol} (Fig. 4a, b), a concentration similar to that described for oligomerization of the enzyme²⁰. Whether 3D^{pol} must form dimers to interact with VPg/poly(A) is not yet known.

3D^{pol} is an exceptional RNA polymerase because it is strictly primer-dependent^{13,21}, as are all known DNA polymerases. A standard assay for 3D^{pol} activity is the transcription of poly(A) in the presence of oligo(U) or oligo(dT) primers^{13,21}. We have used this assay and compared the priming ability of VPg, VPgpU²², and oligo(dT) at different 3D^{pol} concentrations. The oligonucleotide

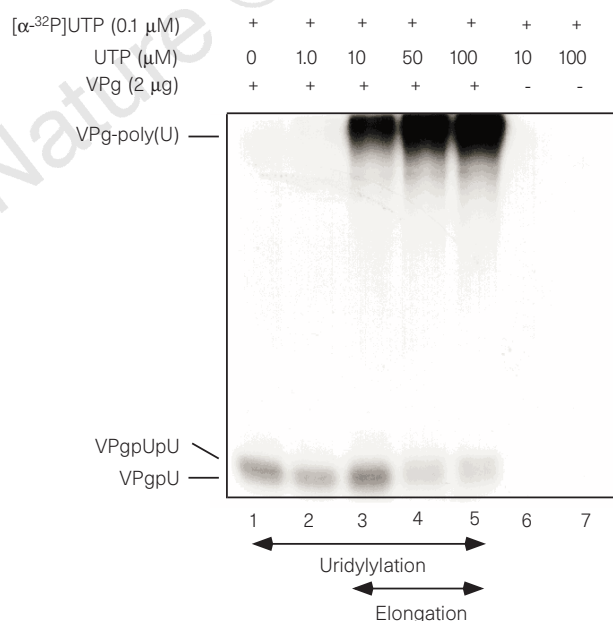


Figure 2 Uridylylation of VPg and synthesis of VPg-poly(U) as a function of UTP concentration. See Methods for details. The VPg concentration was 50 μ M.

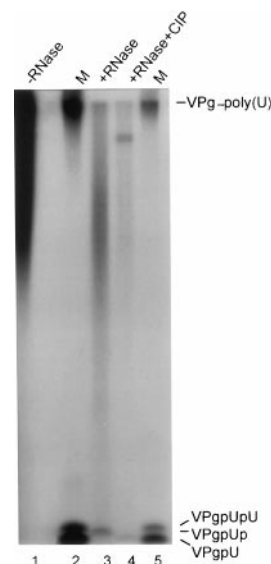


Figure 3 Identification of the elongated product as VPg-poly(U). A 240- μ l reaction (10 μ M UTP, 70 μ Ci[α -³²P]UTP; see Methods), with carrier DNA added, was extracted with phenol/chloroform and was precipitated in ethanol. One-third of the isolated polymeric product remained untreated (lane 1); one-third of the product was treated with RNase A and V₁ (lane 3); and one-third of the product was treated with RNases A and V₁ and CIP (lane 4). Aliquots of untreated reaction mixture (lanes 2, 5) are markers (M).

was far superior to the protein primers (Fig. 4c and Methods). However, all three reactions were optimal at roughly the same enzyme concentration (Fig. 4c). The difference may be due to the high efficiency with which the oligo(dT) primer will bind to any part of the poly(A) template. [³H-uracil]UTP was used in these experiments as a substrate to confirm that the resulting polymer was indeed VPg-poly(U).

We repeated the experiments with a genetic variant of 3D^{pol} to

exclude the possibility that the uridylation/elongation reactions might have been catalysed by contaminating activities that accidentally co-purified with the recombinant 3D^{pol} from *Escherichia coli* cells. A M394 → T mutation in 3D^{pol} confers a temperature-sensitive phenotype to poliovirus replication, specifically to the initiation of synthesis of minus-strand RNA *in vivo* and *in vitro*²³. Purified M394T-3D^{pol} catalysed uridylation and elongation at 30 °C, albeit with reduced activity as compared with

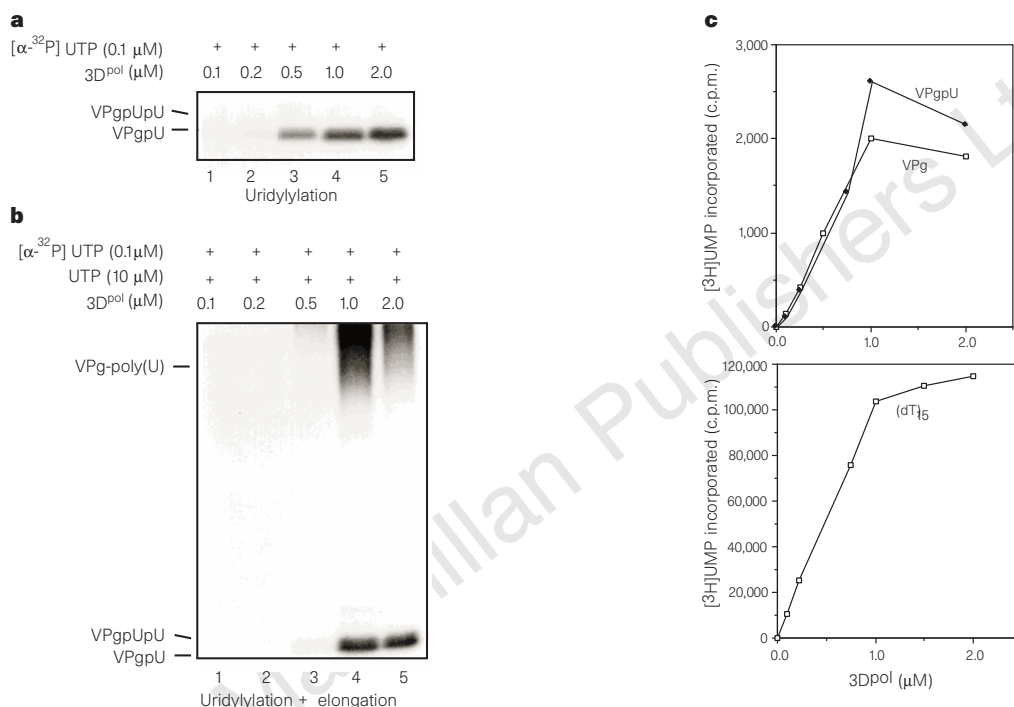


Figure 4 Uridylation of VPg and synthesis of VPg-poly(U) as a function of 3D^{pol} concentration. **a**, Uridylation of VPg. The standard assay was used (see Methods) and the 3D^{pol} concentration was varied as indicated. **b**, VPg-poly(U) synthesis. The standard assay, with 10 μM UTP, was used (see Methods); the

3D^{pol} concentration was varied as indicated. **c**, VPg-poly(U) or poly(U) synthesis with VPg, VPgpU or oligo(dT)₁₅ primers. The filter assay (see Methods) was used in the presence of [³H]UTP/50 μM UTP with VPg or VPgpU as primer (top), or oligo(dT)₁₅ as primer (bottom), as indicated.

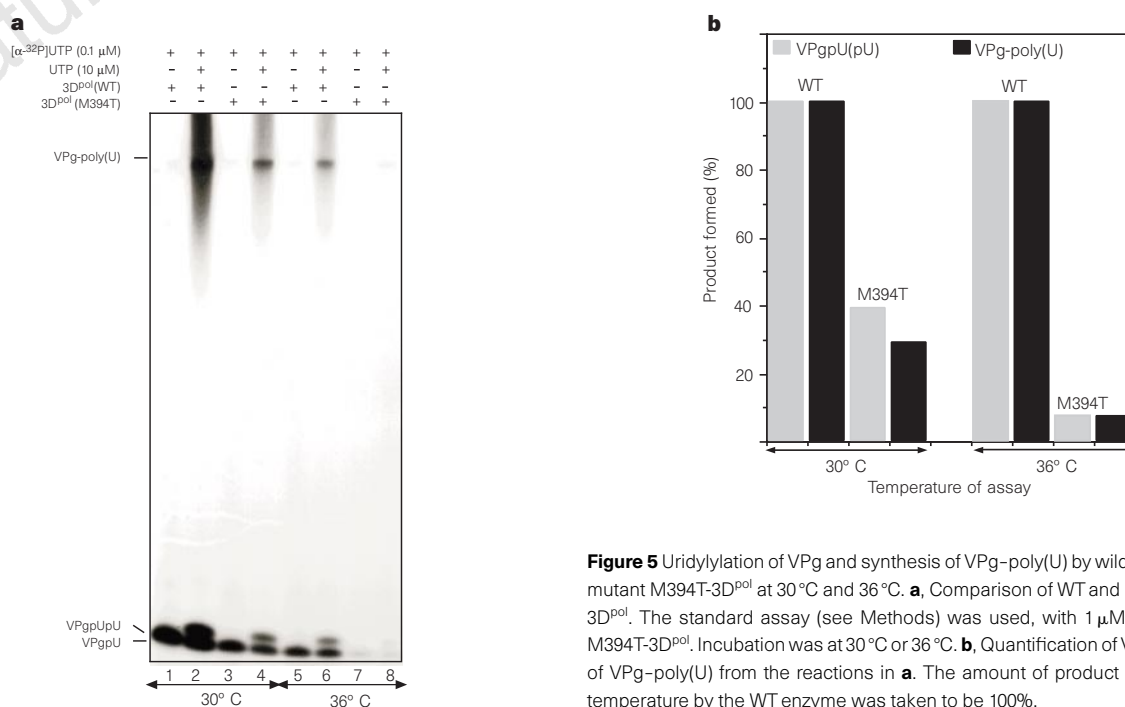


Figure 5 Uridylation of VPg and synthesis of VPg-poly(U) by wild-type (WT) and mutant M394T-3D^{pol} at 30 °C and 36 °C. **a**, Comparison of WT and mutant M394T-3D^{pol}. The standard assay (see Methods) was used, with 1 μM WT or mutant M394T-3D^{pol}. Incubation was at 30 °C or 36 °C. **b**, Quantification of VPgpU(pU) and of VPg-poly(U) from the reactions in **a**. The amount of product made at either temperature by the WT enzyme was taken to be 100%.

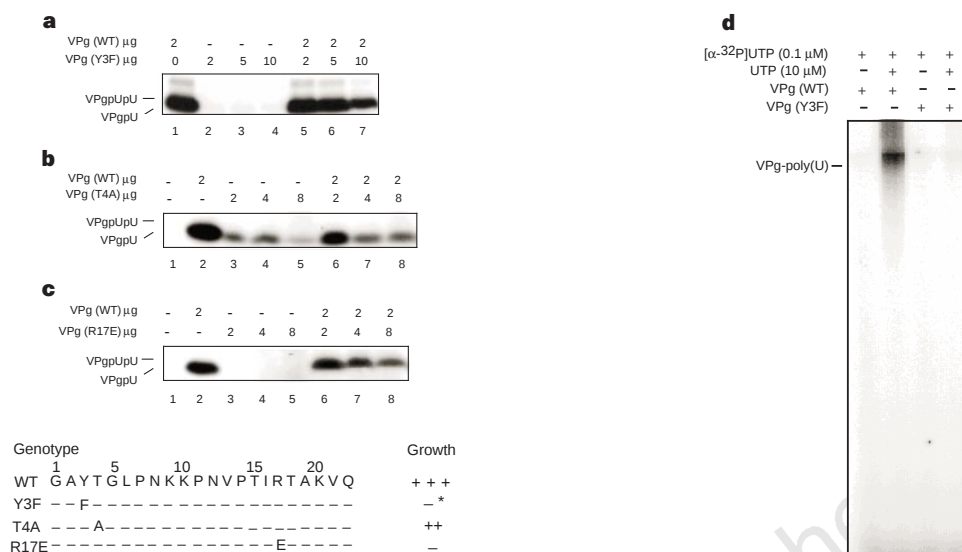


Figure 6 The use of wild-type and mutant VPg peptides as substrates for uridylation by 3D^{pol}. The reaction mixtures (see Methods) contained wild-type (WT) or mutant VPg peptides. **a**, Comparison of WT VPg and Y3F peptides. **b**, Comparison of WT VPg and T4A peptides. **c**, Comparison of WT VPg and R17E peptides. Bottom, the amino-acid sequences of WT, Y3F, T4A and R17E peptides are shown with the viability of engineered genomes containing the WT or mutant VPg peptides²⁶⁻²⁷. The asterix indicates 'quasi-infectious' virus²⁶. **d**, Comparison of WT VPg and Y3F peptides as substrates for uridylation and elongation.

wild-type 3D^{pol} (Fig. 5). At 36 °C, however, M394T-3D^{pol} was essentially inactive (<7% wild-type activity) in both reactions (Fig. 5b). This result indicated that the temperature-sensitive phenotype of the poliovirus M394T-3D^{pol} mutant²³ is related to a defect in uridylation of VPg at the non-permissive temperature. Both uridylation and elongation were completely inhibited by 15 μ M gliotoxin (data not shown), a fungal metabolite shown previously to inhibit both poliovirus RNA replication *in vivo* and oligo(U)-primed poly(U) synthesis by purified 3D^{pol} *in vitro*²⁴. These data show that 3D^{pol} is the activity that catalyses

uridylation of VPg and elongation of the poly(U) chain from VPgU(pU).

Experiments carried out with sequence variants of VPg confirmed that the bond between VPg and nucleotide in VPgU(pU) is O⁴-(5'-uridylyl)tyrosine^{8,9,25,26}. No product was formed with the mutant VPg(Y3F), as was expected (Fig. 6a, lanes 2-4). Moreover, VPg(Y3F) failed to generate VPg-poly(U) (Fig. 6d, lane 4). Elongation depends, therefore, on functioning VPg. We have noticed previously that a poliovirus VPg(T4A) mutant has a growth defect²⁵, a genetic defect that correlated with inefficient uridylation of VPg(T4A) in a standard reaction (Fig. 6b, lanes 3-5). We tested a VPg(R17E) variant, because changes of this arginine residue to different amino acids (R17E, R17Q, R17K) are lethal to poliovirus^{26,27}. VPg(R17E) did not act as a substrate for either uridylation (Fig. 6c, lanes 3-5) or elongation (data not shown). Both non-reactive variants (Y3F and R17E) interfered with the uridylation reaction of wild-type VPg, but only when present at relatively high concentrations (Fig. 6a, c, lanes 5-7 and 6-8, respectively).

Our data describe a mechanism of protein-primed initiation of poliovirus genome replication that explains numerous biochemical and genetic data^{10-12,14-18}. We propose that VPg, 3D^{pol}, poly(A), and UTP form a complex that allows transfer of UMP to the hydroxyl of residue Y3 of VPg (Fig. 7). Transcription of the poly(A) template is then initiated at the 3' hydroxyl group of UMP (Fig. 7, arrow). These results settle a long debate of whether initiation of genome replication occurs by hairpin priming followed by VPg-catalysed processing of covalently linked double-stranded RNA, or whether initiation is due to protein-mediated priming (reviewed in ref. 28). We predict that protein-mediated priming will function in the replication of most, if not all, plus-strand RNA viruses whose genome is protein-linked²⁹, including the *Picornaviridae*, one of which is poliovirus. This mechanism is similar to the initiation of DNA replication of DNA viruses that have linear, double-stranded, protein-linked genomes, such as *Bacillus subtilis* phage ϕ 29 or human adenoviruses^{29,30}. The essential role of residue R17 in VPg

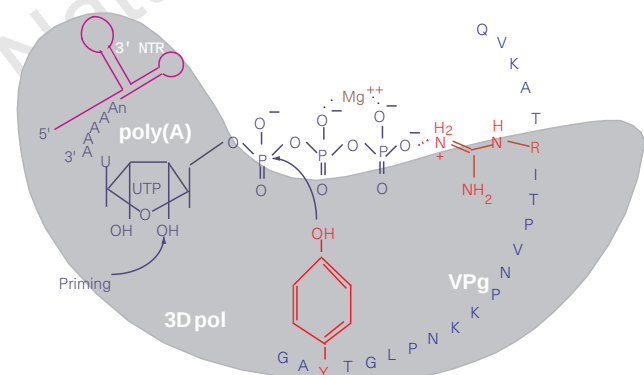


Figure 7 Model of the initiation of poliovirus RNA replication. Polymerase 3D^{pol} in a complex with the 5'-terminal protein VPg binds near the 3'-terminal end (3' NTR) of the poly(A) tail of the poliovirus genome. The complementary nucleotide UTP is selected and positioned near the tyrosine of VPg, a process aided by R17 of VPg. 3D^{pol} catalyzes the formation of a phosphodiester bond between UMP and the hydroxyl group of tyrosine. The 3' hydroxyl group of VPgU then primes synthesis of VPg-poly(U), the 5' end of minus RNA strands. The RNA structures preceding poly(A), depicted as two stem-loops (top left), may be involved in template selection²⁸.

may mirror a similar function of basic amino acids in protein-primed DNA replication of phage $\phi 29$ (ref. 30). This similarity leads to an intriguing question: at what stage during evolution did these diverse viruses copy the basic strategy of genome replication from each other? □

Methods

Assay of uridylylation and of VPg-poly(U) synthesis. Poliovirus polymerase was expressed in *E. coli* from plasmid pT5T-3D (a gift from K. Kirkegaard) and was purified as described²⁰. The enzyme is >99% pure as judged by Coomassie gel staining (data not shown). Purified mutant M394T-3D^{pol} was a gift from J. Lyle and K. Kirkegaard. The chemical synthesis of VPg peptides and of VPgUpU has been described²². [α -³²P]UTP was purchased from either DuPontNEN (3,000 Ci mmol⁻¹) or Amersham (800 Ci mmol⁻¹); [5,6-³H]UTP (43 Ci mmol⁻¹) is an ICN product. RNase V₁ is from Pharmacia; RNase A and CIP (calf intestine alkaline phosphatase) are from Boehringer. Uridylylation of VPg was assayed in a reaction mixture (20 μ l) that contained 50 mM HEPES buffer, pH 7.0–7.5, 3.5 mM magnesium acetate, 6–8% glycerol, 0.5 μ g/0.35 μ M poly(A) (~200 nucleotides long), 2 μ g/50 μ M synthetic VPg, 1 μ g/1 μ M purified 3D^{pol} and 2 μ Ci 0.1 μ M [α -³²P]UTP. To measure VPg-poly(U) synthesis, unlabelled UTP (>10 μ M) was also included in the reaction mixture. Samples were incubated at 33 °C for 60 min and analysed by Tricine-SDS-PAGE (BioRad) with 13.5% polyacrylamide. Gels were dried without fixing and autoradiographed. VPg-poly(U) synthesis was also measured using a DEAE-filter-binding assay²¹ with 1 μ Ci [5,6-³H]UTP replacing [α -³²P]UTP in the standard reaction mixture. When indicated, oligo(dT)₁₅ was used as a primer at 0.05 μ g/0.05 μ M. Background values in this assay with no 3D^{pol} or no VPg were <90 c.p.m.

Quantification of reaction products. Incorporation of ³²P into reaction products was quantified using a phosphorimager (Molecular Dynamics, Storm 860) and incorporation of ³H was quantified by a liquid scintillation analyser. We estimate that in a typical uridylylation/elongation reaction (10 μ M UTP), 1% of the total label is incorporated into each of VPgUpU and VPg-poly(U) and 0.4% into VPgUpU. Of the total VPg added, 0.2% was converted to VPgUpU, 0.04% to VPgUpU, and 0.001% to VPg-poly(U). The molar ratio of UMP incorporated per primer is roughly 30 for oligo(dT)₁₅ and 0.006 for either VPg or VPgUpU.

Received 4 June 1997; accepted 20 February 1998.

- Lee, Y. F., Nomoto, A. & Wimmer, E. In *Progress in Nucleic Acid Research and Molecular Biology* (eds Davidson, J. N. & Coh, W. E.) 89–96 (Academic, New York, 1976).
- Lee, Y. F., Nomoto, A., Detjen, B. M. & Wimmer, E. A protein covalently linked to poliovirus genome RNA. *Proc. Natl Acad. Sci. USA* **74**, 59–63 (1977).
- Flanagan, J. B., Peterson, R. F., Ambros, V., Hewlett, M. J. & Baltimore, D. Covalent linkage of a protein to a defined nucleotide sequence at the 5' terminus of the virion and replicative intermediate RNAs of poliovirus. *Proc. Natl Acad. Sci. USA* **74**, 961–965 (1977).
- Jang, S. K. *et al.* A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during *in vitro* translation. *J. Virol.* **62**, 2636–2643 (1988).
- Pelletier, J. & Sonnenberg, N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**, 320–325 (1988).
- Molla, A., Jang, S. K., Paul, A. V., Reuer, Q. & Wimmer, E. Cardioviral internal ribosomal entry site is functional in a genetically engineered dicistronic poliovirus. *Nature* **356**, 255–257 (1992).
- Kitamura, N. *et al.* Primary structure, gene organization, and polypeptide expression of poliovirus RNA. *Nature* **291**, 547–553 (1981).
- Rothberg, P. G., Harris, T. J. R., Nomoto, A. & Wimmer, E. O⁴-(5'-uridylyl)tyrosine is the bond between the genome-linked protein and the RNA of poliovirus. *Proc. Natl Acad. Sci. USA* **75**, 4868–4872 (1978).
- Ambros, V. & Baltimore, D. Protein is linked to the 5' end of poliovirus RNA by a phosphodiester linkage to tyrosine. *J. Biol. Chem.* **253**, 5263–5266 (1978).
- Takegami, T., Kuhn, R. J., Anderson, C. W. & Wimmer, E. Membrane-dependent uridylylation of the genome-linked protein VPg of poliovirus. *Proc. Natl Acad. Sci. USA* **80**, 7447–7451 (1983).
- Takeda, N., Kuhn, R. J., Yang, C.-E., Takegami, T. & Wimmer, E. Initiation of poliovirus plus-strand RNA synthesis in a membrane complex of infected HeLa cells. *J. Virol.* **60**, 43–53 (1986).
- Toyoda, H., Yang, C.-E., Takeda, N., Nomoto, A. & Wimmer, E. Analysis of RNA synthesis of type 1 poliovirus by using an *in vitro* molecular genetic approach. *J. Virol.* **61**, 2816–2822 (1987).
- Flanagan, J. B. & Baltimore, D. Poliovirus-specific primer-dependent RNA polymerase able to copy poly(A). *Proc. Natl Acad. Sci. USA* **74**, 3677–3680 (1977).
- Yogo, Y. & Wimmer, E. Polyadenylic acid at the 3'-terminus of poliovirus RNA. *Proc. Natl Acad. Sci. USA* **69**, 1877–1882 (1972).
- Yogo, Y. & Wimmer, E. Sequence studies of poliovirus RNA. III. Polyuridylic acid and polyadenylic acid as components of the purified poliovirus replicative intermediate. *J. Mol. Biol.* **92**, 467–477 (1975).
- Dorsch-Haesler, K., Yogo, Y. & Wimmer, E. Replication of picornaviruses. I. Evidence from *in vitro* RNA synthesis that poly(A) of the poliovirus genome is genetically coded. *J. Virol.* **16**, 1512–1517 (1975).
- Nomoto, A., Detjen, B. M., Pozzati, R. & Wimmer, E. The location of the polio genome protein in viral RNAs and its implication for RNA synthesis. *Nature* **268**, 208–213 (1977).
- Wimmer, E. Genome-linked proteins of viruses. *Cell* **28**, 199–201 (1982).

- Esteban, J. A., Bernad, A., Salas, M. & Blanco, L. Metal activation of synthetic and degradative activities of $\phi 29$ DNA polymerase, a model enzyme for protein-primed DNA replication. *Biochemistry* **31**, 350–359 (1992).
- Pata, J. D., Schultz, S. C. & Kirkegaard, K. Functional oligomerization of poliovirus RNA-dependent RNA polymerase. *RNA* **1**, 466–477 (1995).
- Paul, A. V., Cao, X., Harris, K. S., Lama, J. & Wimmer, E. Stimulation of poly(U) synthesis *in vitro* by purified poliovirus protein 3AB. *J. Biol. Chem.* **269**, 29173–29181 (1994).
- Dreef-Tromp, C. M., van den Elst, H., van den Boogart, J. E., van der Marel, G. A. & van Boom, J. H. Solid-phase synthesis of an RNA nucleotide fragment from the nucleoprotein of poliovirus. *Nucleic Acids Res.* **20**, 2435–2439 (1992).
- Barton, D. J. *et al.* Poliovirus RNA polymerase mutation 3D-M394T results in a temperature-sensitive defect in RNA synthesis. *Virology* **217**, 459–469 (1996).
- Rodriguez, P. L. & Carrasco, L. Gliotoxin: inhibitor of poliovirus RNA synthesis that blocks the viral RNA polymerase 3D^{pol}. *J. Virol.* **66**, 1971–1976 (1992).
- Cao, X. & Wimmer, E. Intragenomic complementation of a 3AB mutant in dicistronic polioviruses. *Virology* **209**, 315–326 (1995).
- Kuhn, R. J. *et al.* Construction of a 'mutagenesis cartridge' for poliovirus genome-linked viral protein: isolation and characterization of viable and nonviable mutants. *Proc. Natl Acad. Sci. USA* **85**, 519–523 (1988).
- Xiang, W., Cuconati, A., Paul, A. V., Cao, X. & Wimmer, E. Molecular dissection of the multifunctional poliovirus RNA-binding protein 3AB. *RNA* **1**, 892–904 (1995).
- Richards, O. & Ehrenfeld, E. Poliovirus RNA replication. *Curr. Top. Microbiol. Immunol.* **161**, 89–119 (1990).
- Salas, M. Protein-priming of DNA replication. *Annu. Rev. Biochem.* **60**, 39–71 (1991).
- Hsieh, J.-C., Yoo, S.-K. & Ito, J. An essential arginine residue for protein-primed DNA replication. *Proc. Natl Acad. Sci. USA* **87**, 8665–8669 (1990).

Acknowledgements. We thank M. Salas for helpful discussion; J. Lyle and K. Kirkegaard for purified M394T-3D^{pol}; and A. Wimmer and M. Shepley for critical reading of the manuscript. This work was supported by the National Institute of Allergy and Infectious Diseases of the NIH.

Correspondence and requests for materials should be addressed to A.V.P. (e-mail: apaul@asterix.bio.sunysb.edu).

Distinct roles of the co-activators p300 and CBP in retinoic-acid-induced F9-cell differentiation

Hiroaki Kawasaki[†], Richard Eckner[‡], Tso-Pang Yao[§], Kazunari Taira[†], Robert Chiull^{||}, David M. Livingston[§] & Kazunari K. Yokoyama^{*}

^{*} Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba Science City 305-0074, Japan

[†] Institute of Applied Biochemistry, University of Tsukuba, Tennoudai 1-1-1, Tsukuba Science City 305-0006, Japan

[‡] Institute for Molecular Biology, University of Zürich, CH-8057 Zürich, Switzerland

[§] Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts 02115, USA

^{||} Department of Surgery/Oncology, UCLA School of Medicine and Jonsson Comprehensive Cancer Centre, Los Angeles, California 90024, USA

The related proteins p300 and CBP (cAMP-response-element-binding protein (CREB)-binding protein) are transcriptional co-activators that act with other factors to regulate gene expression^{1–5} and play roles in many cell-differentiation and signal transduction pathways^{6–10}. Both proteins have intrinsic histone-acetyltransferase activity^{11,12} and may act directly on chromatin, of which histone is a component, to facilitate transcription. They are also involved in growth control pathways, as shown by their interaction with the tumour suppressor p53 (refs 13–15) and the viral oncogenes E1A (refs 1, 2, 16) and SV40 T antigen⁵. Here we report functional differences of p300 and CBP *in vivo*. We examined their roles during retinoic-acid-induced differentiation, cell-cycle exit and programmed cell death (apoptosis) of embryonal carcinoma F9 cells^{17–20}, using hammerhead ribozymes capable of cleaving either p300 or CBP messenger RNAs. F9 cells expressing a p300-specific ribozyme became resistant to retinoic-acid-induced differentiation, whereas cells expressing a CBP-specific ribozyme were unaffected. Similarly, retinoic-acid-induced transcriptional upregulation of the cell-cycle inhibitor

p21^{Cip1} required normal levels of p300, but not CBP, whereas the reverse was true for p27^{Kip1}. In contrast, both ribozymes blocked retinoic-acid-induced apoptosis, indicating that both co-activators are required for this process. Thus, despite their similarities, p300 and CBP have distinct functions during retinoic-acid-induced differentiation of F9 cells.

We constructed ribozyme-expression vectors based on transfer RNA-Val targeted to p300 or CBP transcripts²¹. The GUC codons contained within nucleotides 364–382 and 484–502, corresponding to the amino-terminal regions of p300 and CBP, respectively, were selected as cleavage sites (Fig. 1a). To rescue the expression of p300 and CBP, we also constructed a version of p300 or CBP RNA that is not cleavable (nc) by the above ribozymes (p300^{nc} and CBP^{nc},

Fig. 1a). Several independent stable lines of F9 cells were generated using each ribozyme. As F9-cell clones containing the same ribozyme showed comparable phenotypes, one specific clone for each ribozyme was selected for further studies. The concentration of p300 and CBP transcripts from these cell lines was measured by competitive reverse transcription with polymerase chain reaction (RT-PCR; Fig. 1b). The level of p300 mRNA was drastically reduced by about 100-fold in F9 cells expressing an active p300-directed ribozyme (lanes 6–10 in Fig. 1b). In the same cells, the level of CBP transcripts was unchanged (lanes 16–20 in Fig. 1b). An inactive ribozyme targeted to p300 mRNA did not alter mRNA levels for p300 or CBP (lanes 1–5 and 11–15 in Fig. 1b). A similar reduction in the concentration of CBP transcripts was observed in F9 cells

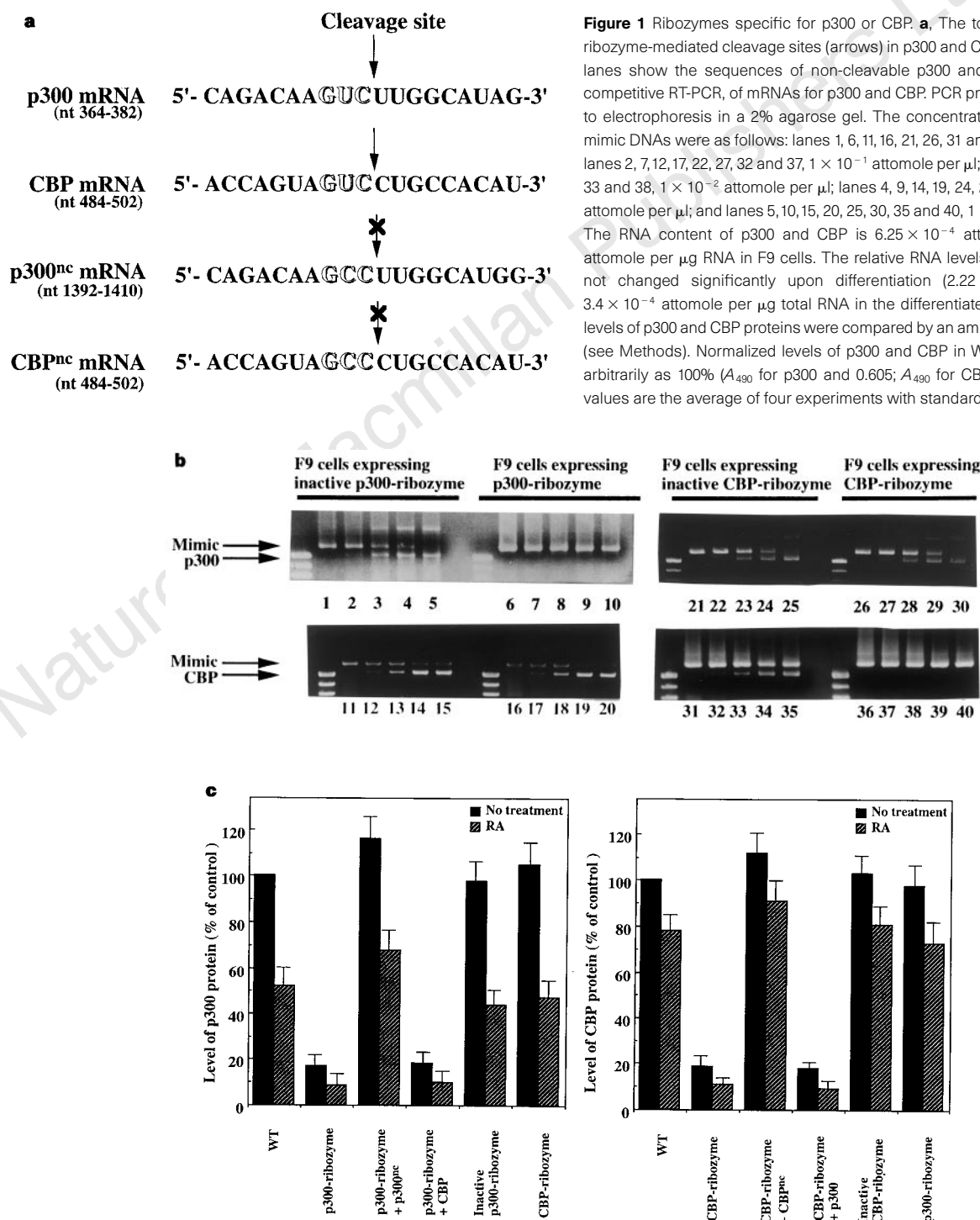


Figure 1 Ribozymes specific for p300 or CBP. **a**, The top two rows show the ribozyme-mediated cleavage sites (arrows) in p300 and CBP and the bottom two lanes show the sequences of non-cleavable p300 and CBP. **b**, Analysis, by competitive RT-PCR, of mRNAs for p300 and CBP. PCR products were subjected to electrophoresis in a 2% agarose gel. The concentrations of p300 and CBP mimic DNAs were as follows: lanes 1, 6, 11, 16, 21, 26, 31 and 36, 1 attomole per μ l; lanes 2, 7, 12, 17, 22, 27, 32 and 37, 1×10^{-1} attomole per μ l; lanes 3, 8, 13, 18, 23, 28, 33 and 38, 1×10^{-2} attomole per μ l; lanes 4, 9, 14, 19, 24, 29, 34 and 39, 1×10^{-3} attomole per μ l; and lanes 5, 10, 15, 20, 25, 30, 35 and 40, 1×10^{-4} attomole per μ l. The RNA content of p300 and CBP is 6.25×10^{-4} attomole and 5.0×10^{-4} attomole per μ g RNA in F9 cells. The relative RNA levels of p300 and CBP are not changed significantly upon differentiation (2.22×10^{-4} attomole and 3.4×10^{-4} attomole per μ g total RNA in the differentiated F9 cells). **c**, Relative levels of p300 and CBP proteins were compared by an amplified sandwich ELISA (see Methods). Normalized levels of p300 and CBP in WT F9 cells were taken arbitrarily as 100% (A_{490} for p300 and 0.605; A_{490} for CBP was 0.71). Indicated values are the average of four experiments with standard deviation.

expressing a CBP-specific ribozyme (lanes 36–40 in Fig. 1b). These data indicate that the ribozymes used act specifically on either p300 or CBP transcripts. Compared with wild-type (WT) F9 cells, p300 and CBP mRNAs were reduced by about 100-fold in the ribozyme-expressing cell lines.

As shown in Fig. 1c, the level of p300 protein in F9 cells expressing p300-directed ribozyme was at least sixfold lower than in F9 cells carrying the inactive p300-directed ribozyme and in WT F9 cells. A comparable, specific reduction of CBP protein levels was seen in cells containing the CBP-directed ribozyme. Importantly, the p300-directed ribozyme did not diminish CBP expression and conversely, the CBP-directed ribozyme did not affect p300 protein levels. As expected, stable introduction of p300^{nc} and CBP^{nc}, both of which are not cleavable by the ribozymes, largely restored p300 or CBP production in the appropriate ribozyme-containing F9-cell clones.

We then examined whether F9 cells lacking p300 or CBP can still respond to retinoic acid (RA). In contrast to control F9 cells, when challenged with RA, F9 cells with the p300-directed ribozyme showed no morphological signs of differentiation (Fig. 2a). These

cells did not downmodulate levels of the cell-surface antigen SSEA-1 (ref. 22) (Fig. 2a, d) and did not induce synthesis of collagen IV (Fig. 2d) and laminin B1 (data not shown) as control cells did^{18,20}. Importantly, stable expression of a non-cleavable p300 transcript (p300^{nc}) in cells containing p300-directed ribozyme fully rescued differentiation (Fig. 2c, d). Unexpectedly, F9 cells carrying the CBP-directed ribozyme readily changed their morphology in the presence of RA (Fig. 2b), showed only residual SSEA-1 staining (Fig. 2b, d) and induced synthesis of collagen IV (Fig. 2d) and laminin B1 (data not shown). These results indicate that p300, but not CBP, plays a critical role in mediating RA-induced differentiation of F9 cells and that CBP and p300 have, in part, distinct functions.

F9-cell differentiation critically depends on the activity of RA receptor/retinoid X receptor (RAR–RXR) heterodimers²³. We thus examined, by transfection assay, the transcriptional activity of RAR–RXR at the promoter of RAR β in cells containing p300- and CBP-directed ribozymes. As shown in Fig. 2e, expression of a RARE β (retinoic acid response element β) promoter–luciferase

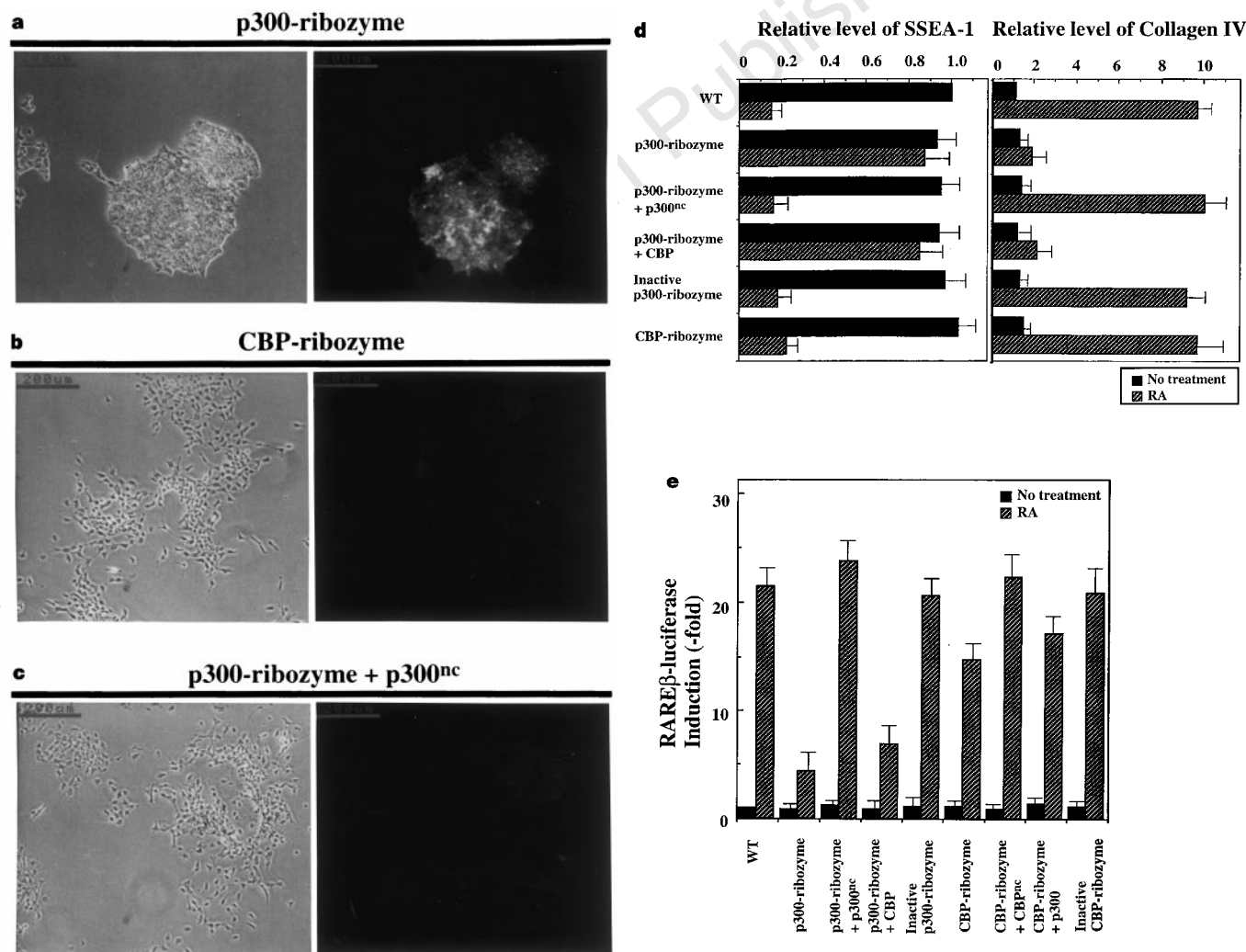


Figure 2 Effects of p300- and CBP-directed ribozymes on F9-cell differentiation and RA-dependent transcription. **a–c**, Left panels show phase-contrast photomicrographs of RA-treated F9 cells containing a p300^{nc} (**a**) or a CBP-ribozyme (**b**), and RA-treated F9 cells expressing both a p300-ribozyme and p300^{nc} (**c**); right panels show immunofluorescence staining of F9 cells with SSEA-1 antibodies. **d**, Left, relative levels of SSEA-1, a marker for undifferentiated cells, after 5 days of treatment with 3×10^{-7} M RA. Right, relative levels of the differentiation marker protein collagen IV in WT F9 cells and cells expressing the indicated ribozyme

after treatment with 3×10^{-7} M RA. Normalized levels of SSEA-1 and collagen IV were taken arbitrarily as 1.0 (A_{490} values for SSEA-1 and collagen IV were 0.8 and 0.08, respectively). All values are the average of four experiments with standard deviation. **e**, Relative luciferase activities of the RARE β -luciferase reporter in the indicated F9-cell lines. Before lysis, cells were incubated with 3×10^{-7} M RA for 18 h. Luciferase activities in WT F9 cells were arbitrarily set at 1.0. The values indicated are the average of three experiments with standard deviation.

construct was strongly impaired in F9 cells containing the p300-directed ribozyme and less so in F9 cells expressing the CBP-directed ribozyme. Ectopic expression of p300^{nc} restored the normal activity of the RARE β promoter. Thus, impaired function of RAR–RXR heterodimers in cells with reduced levels of p300 may, in part, underlie the inability of these cells to differentiate in response to RA.

Apart from differentiating, a fraction of F9 cells treated with RA dies by apoptosis¹⁹. F9 cells carrying p300- or CBP-directed ribozymes were much less sensitive to RA-induced nicking of DNA, a hallmark of apoptosis, than were WT cells or those containing inactive ribozymes (Fig. 3a). The occurrence of DNA fragmentation (Fig. 3b) paralleled the results seen in the TdT-mediated dUTP nick end labelling (TUNEL) assay for apoptosis (Fig. 3a)^{24,25}. Apoptosis was restored in ribozyme-containing cells stably expressing p300 or CBP transcripts that cannot be cleaved (p300^{nc} or CBP^{nc}; Fig. 3a). These results show that RA-induced apoptosis requires expression of both intact p300 and intact CBP.

Treatment of F9 cells with RA decreases the rate of proliferation of F9 cells. In the absence of RA, the basal rate of proliferation of F9 cells expressing the p300- or CBP-directed ribozyme was almost the same as that of WT F9 cells (Fig. 4a). Surprisingly, exposure to RA did not reduce proliferation of F9 cells expressing the p300-directed ribozyme. In cells containing the CBP-directed ribozyme, RA treatment diminished DNA synthesis by only 40–50%, as compared with the 80% inhibition seen in WT F9 cells (Fig. 4a).

To gain insights into the molecular mechanisms underlying the failure of cells carrying ribozymes to reduce proliferation rates, we analysed the expression of inhibitors of cyclin-dependent kinases (CDKs), such as p21^{Cip1} and p27^{Kip1} (ref. 26) (Fig. 4b, c). Expression of p21^{Cip1} is induced during differentiation of many cell types, including F9 cells (Fig. 4b), and after certain events that trigger cell-cycle arrest²⁶. Consistent with the failure of cells containing the p300-directed ribozyme to differentiate and decrease proliferation, p21^{Cip1} levels were not upregulated in these cells after RA treatment. F9 cells expressing the CBP-directed ribozyme induced synthesis of p21^{Cip1} to levels comparable to those in WT F9 cells (Fig. 4b). In striking contrast, induction of p27^{Kip1} synthesis was blocked in F9 cells expressing the CBP-directed ribozyme, whereas levels of p27^{Kip1} were normal in cells expressing the p300-directed ribozyme (Fig. 4c). Introduction of p300^{nc} or CBP^{nc} expression vectors rescued expression of p21^{Cip1} and p27^{Kip1}, respectively (Fig. 4b, c). Unlike the above two CDK inhibitors, the level of the protein p15^{INK4b} did not vary among the various F9 cell lines (data not shown). Thus, expression of p15^{INK4b} appears to be independent of p300 and CBP.

To test whether p300 and CBP regulate p21^{Cip1} and p27^{Kip1} expression at the level of transcription, F9 cells were transfected with reporter plasmids containing the p21^{Cip1} or p27^{Kip1} promoter. In response to RA, the activity of the p21^{Cip1} and p27^{Kip1} promoters was strongly increased (Fig. 4b, c). Concurring with the above data regarding protein levels, the induction of the p21^{Cip1} promoter was specifically dependent on intact p300 levels, whereas induction of p27^{Kip1} required normal CBP levels. The remarkable specificity of p300 but not CBP for the p21^{Cip1} promoter was emphasized by the inability of overexpressed CBP to restore activity of this promoter in cells containing the p300-directed ribozyme. The opposite was true for the p27^{Kip1} promoter (Fig. 4b, c).

Figure 4d shows a schematic representation of the functional roles of p300 and CBP during RA-induced processes. An unexpected finding was that reduced levels of p300 no longer supported differentiation of F9 cells, whereas a comparable decrease in CBP did not compromise differentiation. This observation indicates that p300 and CBP may control, in part, the expression of distinct target genes. Indeed, ribozyme-mediated ablation of p300 expression strongly and specifically affected RA-induced transcription directed by the RARE β and p21^{Cip1} promoters, whereas decreased CBP expression specifically prevented RA-dependent upregulation of

p27^{Kip1} promoter activity. For all three promoters tested, only RA-induced, activated transcription was affected in cells containing p300- or CBP-directed ribozymes; basal transcription was not impaired. This result indicates that p300 and CBP act primarily at the level of induced transcription.

As an independent way to ablate genes, we used antisense phosphorothioate-containing oligodeoxynucleotides (S-ODNs) that were directed against p300 or CBP transcripts. We obtained similar conclusions with respect to the effects of p300 and CBP on cell differentiation, proliferation and apoptosis (data not shown).

Previous reports have indicated that inactivation of one allele of CBP can lead to severe effects in humans (producing Rubinstein-Taybi syndrome (RTS))²⁷ and in mice²⁸. As p300 is presumably wild-type in those patients, it seems that p300 cannot compensate for the loss of one allele of CBP. However, perhaps increasing the expression of p300 in RTS patients may be able to compensate for the diminished CBP expression levels. Our study suggests that this may not be true, as increased CBP expression in cells containing a p300-directed ribozyme did not restore F9-cell differentiation or

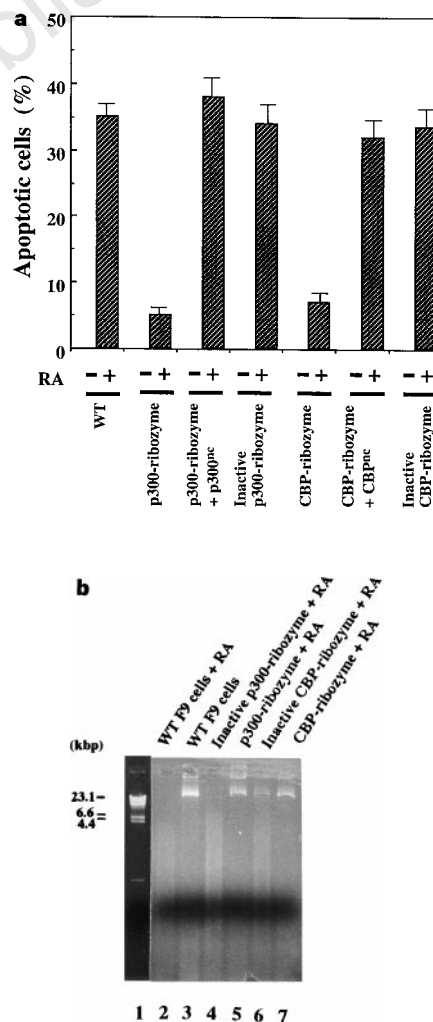


Figure 3 Apoptosis is inhibited in cells containing a p300- or a CBP-directed ribozyme. **a**, Relative percentage of apoptotic cells in cultures of either WT F9 cells or of cells expressing a ribozyme. Cell death was assessed by the TUNEL method²⁴, using cells before and after treatment with RA (3×10^{-7} M). **b**, DNA-fragmentation assay. Nuclei were collected by centrifugation and the chromosomal DNA was analysed by electrophoresis on a 1.5% agarose gel (see Methods). Kbp, kilobase pairs.

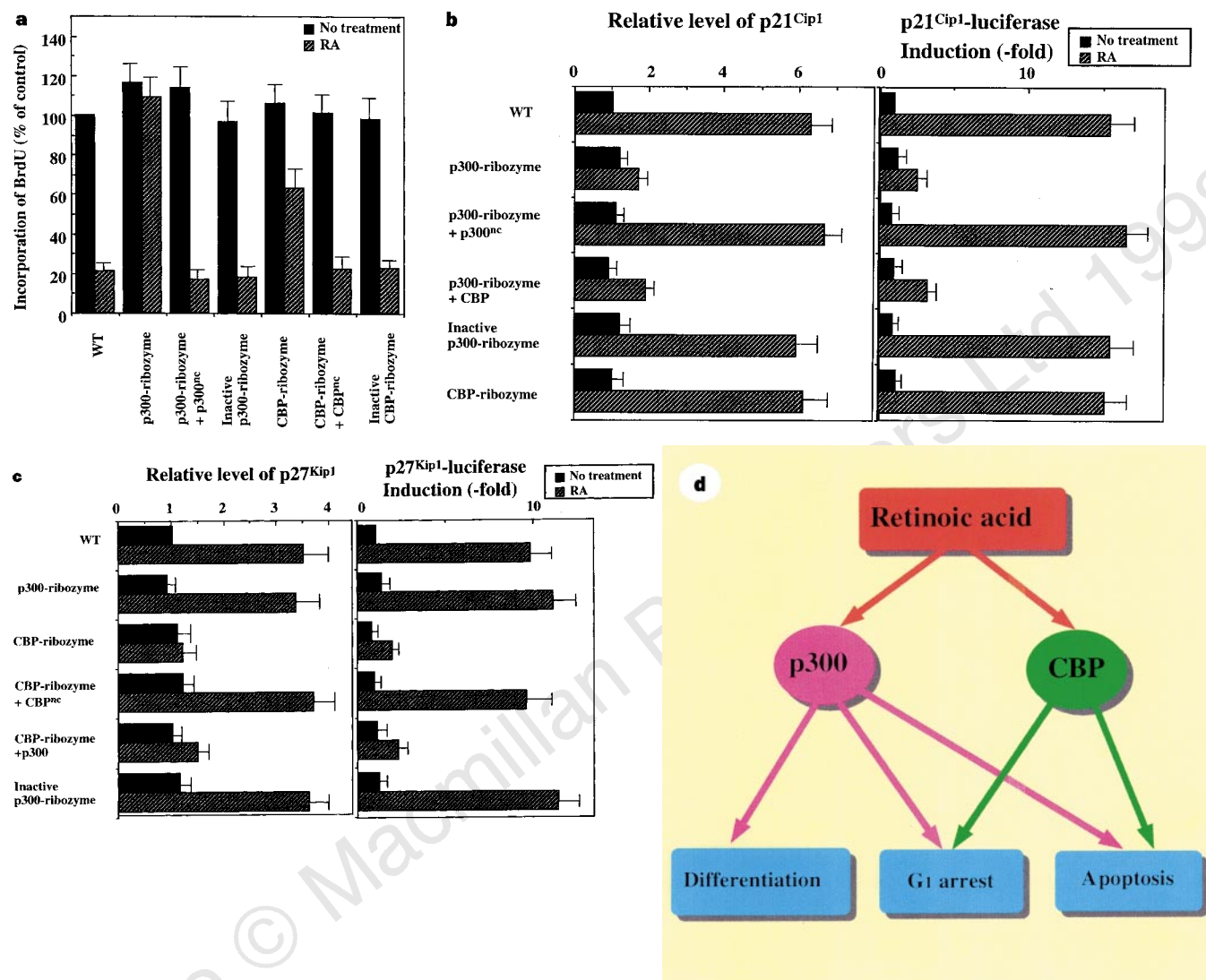


Figure 4 Impact of decreasing p300 or CBP protein levels on cell proliferation. **a**, Relative cell growth of WT or ribozyme-containing F9 cells in the presence or absence of RA (3×10^{-7} M). Incorporation of bromodeoxyuridine (BrdU) was used as a measure of cell growth. Incorporation seen in WT F9 cells without treatment with RA was taken arbitrarily as 100% (A_{405} for F9 cells was 0.79). All values are the average of four experiments with standard deviation. **b**, **c**, Left panels show levels of two CDK inhibitors, p21^{Cip1} (**b**) and p27^{Kip1} (**c**), and right panels show relative luciferase activities of p21-luciferase (WWP-p21-Luc, **b**) and p27-luciferase (p27PL-Luc, **c**) reporters, in the presence or absence of 3×10^{-7} M

RA for 72 h. Normalized levels of CDK inhibitors in F9 cells were taken arbitrarily as 1.0 (A_{490} values for p21^{Cip1} and p27^{Kip1} were 0.12 and 0.2, respectively). All values are the average of four experiments with standard deviation. Luciferase activities of each reporter observed in WT F9 cells without treatment with RA were taken arbitrarily as 1.0. Luciferase activities are the average of three experiments with standard deviation. **d**, Model of RA-induced differentiation. Normal expression of p300, but not CBP, is required for RA-induced differentiation of F9 cells. CBP and p300 are required for both apoptosis and full G1-phase arrest.

RA-mediated induction of p21^{Cip1}, and elevated p300 expression did not rescue p27^{Kip1} transcription in cells containing a CBP-directed ribozyme. Thus, p300 and CBP cannot entirely substitute for each other.

For some RA-dependent responses, such as apoptosis and complete arrest in the G1 phase of the cell cycle, both p300 and CBP are required. As p53 is often involved in these two processes, the recent finding that p300 and CBP are directly linked to the activity of p53 is very intriguing^{13–15}. In particular, a dominant-negative p300 molecule was shown to significantly inhibit both p53-mediated apoptosis and irradiation-induced cell-cycle arrest¹⁴. Thus, the failure to correctly activate p53 may explain in part the resistance of cells containing p300- or CBP-directed ribozymes to apoptosis and full G1-phase arrest.

Our results indicate that, in the context of RA-induced regulation

of cell growth and differentiation, p300 and CBP exert some non-overlapping functions. It will now be important to dissect the precise molecular basis for the surprisingly distinct biological functions of p300 and CBP.

Methods

Plasmids. Construction of the active or inactive ribozyme-expression vectors targeted to p300 and CBP mRNAs, and of pCMVp300 expression vector, has been described^{1,21}. CBP^{nc} and p300^{nc} were generated by substituting the GUC triplets at nucleotide (nt) 1400 and nt 492 of p300 and CBP mRNAs, respectively, with a GCC triplet. p300^{nc} and CBP^{nc} were expressed from the pCEP4 plasmid (Invitrogen, San Diego).

Cells and transfection. WT F9 cells and all stably transformed derivatives thereof were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS) and transfections were carried

out with LipofectAMINE reagent (GIBCO-BRL) as described²¹. Ribozyme-transfected F9 cells were selected by incubation with G418 for 3 weeks. As a control, transformed F9 cells carrying p300- or CBP-directed ribozyme was transfected with 5 µg of the expression vectors for p300^{nc} or CBP^{nc} and selected with hygromycin and G418 for 3 weeks as described^{20,21}.

Competitive RT-PCR analysis. Total RNA was isolated from F9 cells with an Isogen kit (Nippon Gene, Toyama, Japan) according to the manufacturer's protocol. RT-PCR was performed using an RNA PCR Kit version 2 (Takara, Kyoto, Japan) with p300 upstream (nt 173–196) and downstream (nt653–677) primers, or CBP upstream (nt 442–467) and downstream (nt 922–946) primers. The 'mimic' DNA fragments (*v-erbB*) for p300 or CBP were constructed with a PCR MIMIC construction kit (Clontech, Palo Alto, CA) using the primers described above. The p300 mimic DNA and p300 complementary DNA (or CBP mimic DNA and CBP cDNA) were allowed to compete for amplification in the same reaction. By calculating the amount of p300 mimic DNA (or CBP mimic DNA) in a reaction, we determined the amounts of p300 (or CBP) mRNA.

Amplified sandwich enzyme-linked immunosorbent assay (ELISA). Amplified sandwich ELISA was carried out with ELAST ELISA Amplification system (Du Pont, Boston, MA) according to the manufacturer's protocol. The following antibodies were used: monoclonal antibody against SSEA-1 (Kyowa Medex Co., Tokyo); polyclonal antibody against collagen IV (Cosmo Bio., Tokyo); polyclonal antibody against p21^{Cip1} and monoclonal antibody against p27^{Kip1} (Santa Cruz Biotechnology, CA); and monoclonal antibodies against p300 (RW128) (ref. 1) and against CBP (AC238) (ref. 5).

Immunostaining. Cells were fixed in 4% paraformaldehyde in PBS for 1 h. Then cells were incubated with monoclonal antibody against SSEA-1 for 2 h. Fluorescein-isothiocyanate-conjugated (FITC-conjugated) secondary antibodies were then added at a dilution of 1:200 and stained.

Detection of apoptosis and cell growth. The percentage of apoptotic cells was determined by TUNEL analysis²⁴ according to the manufacturer's protocol (Boehringer). The DNA-fragmentation assay was carried out as described²⁵. Cell growth was monitored with a kit for labelling and detection of bromodeoxyuridine as described in the manual (Boehringer).

Assay of luciferase activity. LipofectAMINE reagent (GIBCO-BRL) was used to transiently transfect F9 cells with the indicated luciferase reporter plasmids and with pCMVβ-gal expression plasmid as an internal standard. The transfection medium was replaced by DMEM supplemented with 10% FCS 12 h after transfection. After incubation for 30 h, cells were collected, lysed and assayed for luciferase activity, as described²¹. Luciferase activities were normalized relative to β-galactosidase activities.

Received 10 November 1997; accepted 12 February 1998.

1. Eckner, R. *et al.* Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kDa protein (p300) reveals a protein with properties of a transcriptional adapter. *Genes Dev.* **8**, 869–884 (1994).

2. Lundblad, J. R., Kwok, R. P. S., Laurence, M. E., Harter, M. L. & Goodman, G. H. Adenovirus E1A-associated protein p300 as a functional homologue of the transcriptional co-activator CBP. *Nature* **374**, 85–88 (1995).
3. Arany, Z., Newsome, D., Oldread, E., Livingston, D. M. & Eckner, R. A family of transcriptional adapter proteins targeted by the E1A oncoprotein. *Nature* **374**, 81–84 (1995).
4. Kwok, R. P. S. *et al.* Nuclear protein CBP is a coactivator for the transcription factor CREB. *Nature* **370**, 223–226 (1994).
5. Eckner, R. *et al.* Association of p300 and CBP with Simian virus 40 large T antigen. *Mol. Cell. Biol.* **16**, 3454–3464 (1996).
6. Eckner, R., Yao, T.-P., Oldread, E. & Livingston, D. M. Interaction and functional collaboration of p300/CBP and bHLH proteins in muscle and B-cells differentiation. *Genes Dev.* **10**, 2478–2490 (1996).
7. Kamei, Y. *et al.* A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell* **85**, 403–414 (1996).
8. Chakravarti, D. *et al.* Role of CBP/p300 in nuclear receptor signalling. *Nature* **383**, 99–103 (1996).
9. Yao, T.-P., Ku, G., Zhou, N., Scully, R. & Livingston, D. M. The nuclear hormone receptor coactivator SRC-1 is a specific target of p300. *Proc. Natl Acad. Sci. USA* **93**, 10626–10631 (1996).
10. Nakajima, T. *et al.* RNA helicase A mediates association of CBP with RNA polymerase II. *Cell* **90**, 1107–1112 (1997).
11. Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87**, 953–959 (1996).
12. Bannister, A. J. & Kouzarides, T. The CBP co-activator is a histone acetyltransferase. *Nature* **384**, 641–643 (1996).
13. Gu, W., Shi, X.-L. & Roeder, R. G. Synergistic activation by CBP and p53. *Nature* **387**, 819–823 (1997).
14. Avantaggiati, M. L. *et al.* Recruitment of p300/CBP in p53-dependent signal pathways. *Cell* **89**, 1175–1184 (1997).
15. Lill, N. L., Grossman, S. R., Ginsberg, D., DeCaprio, J. A. & Livingston, D. M. Binding and modulation of p53 by p300/CBP activators. *Nature* **387**, 823–827 (1997).
16. Moran, E. DNA tumor viruses and the cell cycle. *Curr. Opin. Genet. Dev.* **3**, 63–70 (1993).
17. Bernstine, E. G. *et al.* Alkaline phosphatase activity of differentiation in mouse teratoma. *Proc. Natl Acad. Sci. USA* **70**, 3899–3903 (1973).
18. Strickland, S. & Mahdavi, V. The induction of differentiation in teratocarcinoma stem cells by retinoic acid. *Cell* **15**, 393–403 (1978).
19. Atencia, R. *et al.* Apoptosis during retinoic acid-induced differentiation of F9 embryonal carcinoma cells. *Exp. Cell Res.* **214**, 663–667 (1994).
20. Kitabayashi, I. *et al.* Phosphorylation of the adenovirus E1A-associated 300-kDa protein in response to retinoic acid and E1A during differentiation of F9 cells. *EMBO J.* **14**, 3496–3509 (1995).
21. Kawasaki, H. *et al.* Selection of the best target site for ribozyme-mediated cleavage within a fusion gene for adenovirus E1A-associated 300-kD protein (p300) and luciferase. *Nucleic Acids Res.* **24**, 3010–3016 (1996).
22. Solter, D. & Knowles, B. Monoclonal antibody defining a stage-specific mouse embryonic antigen (SSEA-1). *Proc. Natl Acad. Sci. USA* **75**, 5565–5569 (1978).
23. Chambon, P. A decade of molecular biology of retinoic acid receptors. *FASEB J.* **10**, 1187–1197 (1996).
24. Gavrieli, Y., Sherman, Y. & Ben-Sasson, S. A. Identification of programmed cell death *in situ* via specific labeling of nuclear DNA fragmentation. *J. Cell Biol.* **119**, 493–501 (1992).
25. Wyllie, A. H. Glucocorticoid-induced thymocyte apoptosis is associated with endogenous endonuclease activation. *Nature* **284**, 555–556 (1980).
26. Sherr, C. J. & Roberts, J. M. Inhibitors of mammalian G1 cyclin-dependent kinases. *Genes Dev.* **9**, 1149–1163 (1995).
27. Petrij, F. *et al.* Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature* **376**, 348–351 (1995).
28. Tanaka, Y. *et al.* Abnormal skeletal patterning in embryos lacking a single *Cbp* allele: a partial similarity with Rubinstein-Taybi syndrome. *Proc. Natl Acad. Sci. USA* **94**, 10125–10220 (1997).

Acknowledgements. We thank G. H. Goodman, K. Umezono, B. Vogelstein and T. Sakai for gifts of the pCMV-CBP, RAREβ-luc, p21-luc and p27-luc plasmids, respectively; and K. Itakura and G. Gachelin for discussion. T.-P.Y. was supported by the Damon Runyon Walter Winchell Cancer Research Fund. This work was supported by a START grant from the Swiss National Science Foundation (to R.E.), the Life Science Research Project of RIKEN, the Special Cooperation Funds of the Science and Technology Agency, and Grants-in-Aid from the Ministry of Education, Science and Culture (Japan) (to K.K.Y.).

Correspondence and requests for materials should be addressed to K.K.Y. (e-mail: kazunari@rtc.riken.go.jp).